

PROCEEDINGS OF THE
INDIAN ACADEMY OF SCIENCES
(Mathematical Sciences)

VOLUME 90, JANUARY-DECEMBER 1981

INDEX

THE INDIAN ACADEMY OF SCIENCES
BANGALORE 560080

Proceedings of the Indian Academy of Sciences

Vol. 90, January - December 1981

CONTENTS (Mathematical Sciences)

On some problems in singularity theory	<i>V Arnold</i>	1
Yang-Mills and bundles over algebraic curves	<i>M F Atiyah and R Bott</i>	11 ✓
Vanishing theorems for square-integrable harmonic forms	<i>Jozef Dodziuk</i>	21
On operators of trace class in $L^2(X, \mu)$	<i>J J Duistermaat</i>	29
Deformations of metrics and associated harmonic maps	<i>J Eells and L Lemaire</i>	33
Curvature and the heat equation for the de Rham complex	<i>Peter B Gilkey</i>	47 ✓
Units of Hill curves	<i>H P McKean</i>	81
A submersion principle and its applications	<i>Harish Chandra</i>	95
Geometric construction of cohomology for arithmetic groups I	<i>John J Millson and M S Raghunathan</i>	103
Polarisations on an abelian variety	<i>M S Narasimhan and M V Nori</i>	125
Poisson formulae of Hecke type	<i>S Raghavan and S S Rangachari</i>	129
Orthogonal and spin bundles over hyperelliptic curves	<i>S Ramanan</i>	151 ✓
EOL and ETOL array languages	<i>Nalinakshi Nirmal and Kamala Krithivasan</i>	167
On the relation of generalized Valiron summability to Cesàro summability	<i>V K Krishnan</i>	181
On the mean square value of Hurwitz zeta function	<i>M J Narlikar</i>	195
On a generalization of the class of functions with bounded Mocanu variation	<i>J Thangamani</i>	213
Analyse asymptotique des équations de Transport dans le cas d'évolution	<i>Rafael F Santos</i>	219
Application of Newton's method to a homogenization problem	<i>S Kesavan</i>	229 ✓
Homogenization of eigenvalue problems in perforated domains	<i>M Vanninathan</i>	239
A note on the mean value of L-series	<i>V V Rane</i>	273

On some problems in singularity theory

By

V ARNOLD

Much progress in singularity theory of differentiable maps is based on empirical data. Some of these empirical facts later become theorems. We discuss here some facts, that are not explained today and some conjectures, related to them.

1. Semicontinuity of singularity spectrum

Let $f : (C^n, 0) \rightarrow (C, 0)$ be a holomorphic function-germ of finite multiplicity μ . One can associate to such a germ a set of μ rationals (not necessarily all different), which we shall call *singularity spectrum*.

Following Steenbrink [14], we denote spectrum points l_k , $k = 1, \dots, \mu$. The singularity spectrum has the following properties.

(1) Eigenvalues λ_k of the monodromy are related to the spectrum by exponentiation : $\lambda_k = \exp(2\pi i l_k)$.

(2) Let f be quasi-homogeneous, and let $\{x^m\}$ be a monomial C -basis of the local algebra

$$Q_f = C[[x_1, \dots, x_n]]/(\partial f/\partial x_1, \dots, \partial f/\partial x_n).$$

Then $\{l_k\}$ is the set of weights of μ differential forms $x^{m_k} dx_1 \dots dx_n$ (here $\deg x_k = \deg dx_k = a_k$, $\deg f = 1$).

(3) Let f be a function in two variables, which is generic among functions with a given Newton diagram Γ (figure 1).

Then the spectrum consists of orders (for the Newton filtration) of monomials, whose exponents are detectable from figure 1 (the order of a monomial x^m in the Newton filtration is the coefficient λ , for which $\lambda m \in \Gamma$).

(4) In the general case $l_k \bmod 1$ is defined by 1, and the integer part of l_k — by the Steenbrink conventions :

$[l_k] = q$ if l_k corresponds to an eigenvalue of the monodromy on the space $H^{p,q}$ of the mixed Hodge structure on the vanishing cohomology group.

(5) The spectrum is symmetric, with centre $l = n/2$.

Many examples led to the following conjecture. Let the spectra of singularities be ordered : $l_1 \leq l_2 \leq \dots \leq l_\mu$.

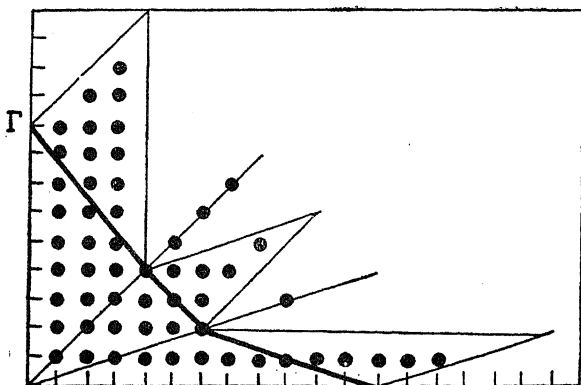


Figure 1. The spectrum of a Newton diagram.

Conjecture. The spectrum is semicontinuous in the following sense: let a singularity S be adjacent to a (simpler) singularity S' (with $\mu' < \mu$), then $l_k \leq l'_k$ for $k = 1, \dots, \mu'$.

Remarks. (1) Even in simple and explicitly calculable cases, like (2) and (3) above, this conjecture is a nontrivial arithmetical assertion on integer points in convex polyhedra.

(2) The conjecture implies the semicontinuity of dimensions of spaces of the Hodge filtration and of the filtration dual to the Hodge filtration in vanishing cohomology, that is the semicontinuity of numbers

$$h^r = \sum_{p \leq r} \sum_a h^{p,a} \quad \text{and} \quad h_r = \sum_{p > r} \sum_a h^{p,a}.$$

(3) In particular, for the case of functions in two variables, these semicontinuity are reduced to the semicontinuity of genus g of the Riemann surface (of the "vanishing cycles manifold") and of the "cogenus" $\mu - g$. Semicontinuity of both numbers g and $\mu - g$ is clear (the last, as Varchenko has explained to the present author, reduced to the fact that the inclusion of the vanishing cycles manifold of a simpler singularity, to that of a more complicated one is monomorphical at the homology level).

(4) The semicontinuity of l_1 , that is of the first spectrum point, is very important for the theory of integral asymptotics. Probably, our conjecture on the semicontinuity of l_1 implies (or is equivalent to) the conjecture on the semicontinuity of oscillating integral asymptotics, which originated from [1], was disproved by Varchenko in [16] and was finally reformulated by Pham [13].

(5) The relation of the spectrum to the set of zeros of Bernstein polynomial (see [3], [12]) seems less proved than one should like, but our conjecture can be reformulated in this setting too (?).

(6) It follows from the spectrum symmetry, that the conjecture implies the two-sided inequalities

$$l_k \leq l'_k \leq l_{k+(\mu-\mu')}.$$

For instance, if only one point bifurcates from a complicated singularity (S) , so that $\mu = \mu' + 1$, then the deformed singularity (S') has a spectrum which divides the spectrum of (S) .

The relation between spectra of (S) and (S') is the same as between semiaxes of an ellipsoid in R^μ and of its section by $R^{\mu'}$. Does there exist a quadratic form, associated with the singularity, whose eigenvalues (in some Euclidean space) are values of a monotonic function at spectrum points?

(7) We can deform the loop, defining the monodromy in the base space of the versal deformation of a complicated singularity into a product of the loop, defining the monodromy of the simpler one and of some "positive" loops, along which the discriminant argument increases. One can conjecture that the image of the product in the matrix group "rotates more than for the simpler monodromy". This gives some heuristic explanation for the semicontinuity. While these ideas are not at all clear, they are sometimes useful; for instance it was precisely these ideas that have led the present author to the conjecture on the formula for the quasi-homogeneous singularity signature (see [6]). In this case, the "rotation" was defined in terms of two-dimensional invariant planes of the symplectic mapping

$$\begin{pmatrix} 0 & \text{Var} \\ -(\text{Var}^t)^{-1} & 0 \end{pmatrix}$$

in $H^* + H_*$. The conjecture was related to the positivity of some eigenvalues of this symplectic mapping in the sense of Krein's parametrical resonance theory. However the proof of the conjecture, given later by Steenbrink [15], is quite different.

2. Bifurcation diagrams of complex singularities

Bifurcation diagrams of real functions at critical points of series A ($A_2 = x^3 + y^2$, $A_3 = x^4 + y^2$, ...) are very useful for the calculations (and definitions) of generalised Whitehead groups in algebraic K -theory (Cerf [5], Hatcher [9], Wagoner [18], Volodin [17]).

This led to the question, what is the "complex analogue" of these algebraical objects? Such "complex analogues" are perhaps quite different from K -theory. For instance, complex analogue for Morse theory is Picard-Lefschetz theory, but it would be very difficult to reconstruct the second theory, knowing nothing but the first. We also know that the complex analogue for $K(\pi, 0)$ is $K(\pi, 1)$, and for the symmetric group—the braids group. The complex analogue for "boundary" is "two-fold-ramified covering". But we have no general methods or axioms for finding such analogues.

As one of the candidates, arising from the complex bifurcation diagrams, we describe a "quasi-resolvent" of the fundamental group of the complement to the singularity bifurcation diagram.

Let Γ_0 be a group, presented as the quotient F_0/R_0 , where F_0 is free and R_0 is an invariant subgroup, generated (as an invariant subgroup) by elements of the form $(af)f^{-1}$, where $f \in F_0$ and $a \in \text{Aut } F_0$, group of automorphisms of F_0 .

Let Γ_1 be a group of automorphisms of F_0 , leaving invariant every class fR_0 , and large enough to generate R_0 . Let us suppose, that Γ_1 is represented as $\Gamma_1 = F_1/R_1$, and so on: we have a chain of groups $\Gamma_i = F_i/R_i$. We call such a chain *quasi-resolvent* of Γ_0 .

Something of this kind arises from the fundamental group Γ_0 of the complement $C^\mu - \Sigma$ to the bifurcations diagram Σ in the versal deformation base space C^μ of a singularity. Let F_0 be $\pi_1(C^1 \setminus \Sigma)$, where C^1 is a generic line in C^μ . Let us consider in C^μ/C^1 the set Σ_1 of non-generic (with respect to Σ) lines. The group

$$\Gamma_1 = \pi_1((C^\mu/C^1) \setminus \Sigma_1)$$

acts on $\pi_1(C^1 \setminus \Sigma)$, leaving invariant elements of $\pi_1(C^\mu \setminus \Sigma)$. Choose in C^μ/C^1 a generic line (that is, a generic C^2 containing C^1 in C^μ). We obtain a set of generators for Γ_1 , that is we consider the free group

$$F_1 = \pi_1((C^2/C^1) \setminus \Sigma_1),$$

and so on, finishing at $F_{\mu-1}$.

In the case of A_μ singularities, the group Γ_0 is the Artin braids group with $\mu + 1$ strings, the free group F_0 is generated by the μ standard generators of the braids group, the group Γ_1 can be considered as a "quasi-relations" group for Γ_0 (not to be confused with R_0). In the same way Γ_2 corresponds to the "quasi-relations between quasi-relations" and so on. But even in the A_μ case it is not clear whether Γ_1 coincides with the whole group of automorphisms of F_0 , which conserves all elements of Γ_0 (and which belongs to the group of automorphisms, generated by orientation-preserving plane with holes homeomorphisms).

Perhaps for the study of $\{\Gamma_i\}$ the classification of all decompositions of simple singularities A, D, E into simpler ones will be useful. Such a classification for functions (not just for levels, which is much easier) is recently found by Ljashko [10].

Very little is known on the topology of the complements to more complicated singularities bifurcations sets. Looijenga [11] has reported that the complements are $K(\pi, 1)$ for parabolical singularities, but his arguments are not clear.

3. Cohomology of complements to bifurcation diagrams

The imbedding of the versal deformation of a simpler singularity (S') base space $C^{\mu'}$ into the base space for a more complicated singularity (S) defines a cohomology homomorphism:

$$H^*(C^\mu \setminus \Sigma) \leftarrow H^*(C^{\mu'} \setminus \Sigma') \quad (\mu > \mu'),$$

between the cohomologies of complements to bifurcation diagrams.

A question naturally arises whether these homomorphisms are canonical and whether one can define a stable cohomology ring (which is, in a sense, the ring of cohomologies of the complement to the bifurcations diagram for $f \equiv 0$ in the infinite-dimensional versal deformation space). Even if this programme cannot be completely realised, one still can associate stable cohomology classes

to (at least) some strata of the natural stratification of the set of function-germs set (or hypersurface-germs set).

Since the cohomology classes of complements to bifurcation diagrams for versal deformations define corresponding classes in the base spaces of arbitrary deformation complement to bifurcation sets, one can hope that any information on the "stable ring" can be useful to obtain information on global properties of bifurcation sets for arbitrary families of functions (hypersurfaces, mappings, ...).

4. Modality

The modality (moduli number) of a Lie group action at a point of a manifold is the minimal integer m for which orbits in some neighbourhood of the point can be arranged in a finite number of families with $\leq m$ parameters.

Problem. Let $f: (R^n, 0) \rightarrow (R, 0)$ be a finite multiplicity real function-germ. Is its modulus number in the real jets space equal to its modulus number in the complex jets space? [The corresponding groups are the groups of jets $(R^n, 0) \rightarrow (R^n, 0)$ and $(C^n, 0) \rightarrow (C^n, 0)$ acting as "right equivalence" $(gf)(x) = f(g(x))$].

I was told by Professor E. B. Vinberg that there exist real representations of real Lie groups, such that the modality of their complexification is larger than the modality of the initial real action. But it is unknown whether such a case is possible for the right equivalence action of the diffeomorphism group on the functions space.

For quasi-homogeneous singularities there exists a notion of "inner modality" which can be calculated as the number of monomials of positive degrees in the monomial basis of the corresponding versal deformations module

$$T_f = (C[x_1, \dots, x_n])^p / \left(\sum_{j=1}^p \frac{\partial f_j}{\partial x_i} \frac{\partial}{\partial y_j}, f_r \frac{\partial}{\partial y_q} \right), \quad i = 1, \dots, n;$$

$$r, q = 1, \dots, p.$$

Here $f_j \in C[x_1, \dots, x_n]$ are quasi-homogeneous polynomials of degrees D_j , where $\deg x_i = A_i$; the module $(C[x])^p$ is generated by p free generators $\partial/\partial y_q$, whose degrees are $-D_q$; one supposes that $p \leq n$ and that $\dim_C T < \infty$.

One conjecture that the inner modality for $p < n$ is equal to the modality (of the contact group action), but it is not proved (as it is a known problem for the case of right-equivalence of functions). The above conjecture is partially confirmed by the theorem (due to I. G. Scherbak) that inner modality 0 complete intersections (quasi-homogeneous curves) in C^3 coincide with contact modality 0 complete intersections (which are all quasi-homogeneous). The classification of these as a standard exercise on Newton diagrams, was at Moscow a known examination problem (1973). By the way, this classification disproves some of the classification results in the sixth part of J. Mather's celebrated paper on singularities.

The Scherbak theorem (containing also quasi-homogeneous unimodular singularities list) was proved with the help of the following formula for the Poincaré polynomial of the graded module T , for $p = n - 1$ (that is, for singularities of generic quasi-homogeneous curves) :

$$p(t) = \frac{\prod (1 - t^{D_j})}{\prod (1 - t^{A_i})} (\sum t^{-D_j} - \sum t^{-A_i} + 1) + t^{\sum D_j - \sum A_i}.$$

One can rewrite this as

$$(*) \quad p(t) = t^{\sum D_j - \sum A_i} h(1/t)$$

denoting by h the Poincaré polynomial of the relative differential forms graded module H -, calculated by Hamm [7]. The formula (*) was found empirically for $n = 3$, $p = 2$. It seems that Ljashko can prove it, but his proof does not explain the duality between modules T and H for quasi-homogeneous curves singularities.

In the more general case of generic complete intersections of positive dimension the Ljashko's formula for the Poincaré polynomial of T , is

$$p(t) = (-1)^{n-p}$$

$$t^{\sum D_j - \sum A_i} \left[-1 + \operatorname{res}_{s=0} \prod \frac{1 - st^{A_i}}{s - st^{A_i}} \prod \frac{s - st^{D_j}}{1 - st^{D_j}} \frac{ds}{1 - s} \right] + \frac{\prod (1 - t^{D_j})}{\prod (1 - t^{A_i})} [\sum t^{-D_j} - \sum t^{-A_i} + 1].$$

Recently Hamm and Gruel have proved $p(1) = h(1)$ for $n - p > 0$, but they give no formula for $p(t)$.

5. Real singularities topology

Mixed Hodge structure defines for every singularity of a function a large set of integers $h_{\lambda}^{p,q}$ (where λ are monodromy eigenvalues). It seems that these numbers are closely related to the real geometry of the function, its level sets and its morsifications and bifurcations diagrams topology.

A simple example of this is the generalised Petrovski inequality

$$|\operatorname{ind}| \leq h_1^{n/2, n/2},$$

for the local degree of the gradient mapping of a real smooth function-germ $f : (R^n, 0) \rightarrow (R, 0)$, n even : see [1] for more details.

Other geometrical invariants of the real singularity, which would be interesting to compare with Hodge numbers, are, for instance :

- (i) Betti numbers of real non-singular neighbour level set (or their partial sums, may be alternate);
- (ii) Numbers of critical points of different indexes, arising from different real morsifications (or their partial sums, may be alternate);

(iii) Numbers describing the possible complication of one real level set of a morsification, for instance, the maximal number of singular points on the same level set.

For empirical works in this direction it is possible to use the Steenbrink conjecture [14] (while this conjecture is wrong as it is stated). A comparison of Hodge numbers, calculated from the Newton diagram by formulas of Steenbrink, Danilov and Kirillov, with the values of geometrical invariants, like those above will, perhaps, generate new (and best possible) inequalities for the real geometry invariants.

However even the Petrovski inequality is known to be best possible only in the simplest cases. For example, it is not known how large the Poincaré index (= the local degree of the gradient map) can be for a real homogeneous function f of degree m in n variables, $f: R^n \rightarrow R$.

We only know that the Petrovski inequality gives the exact maximum for $n = 3$ (that is, for curves in RP^3) or for $m = 3$ [the extremal function $f = (x_1 + \dots + x_n)^3 - x_1^3 - \dots - x_n^3$ was constructed by D B Fuks]. For $m = n = 4$ (surfaces of degree 4 in RP^3) the inequality is the best possible too.

The classification of real projective surface of degree 4 in RP^3 was one of the questions in the 16th Hilbert problem, at present this classification is known completely, after the works of V M Harlamov (see [8]) and V V Nikulin; they find not only the topological types, but also all possible isotopical types in RP^3 and even classify components of the complement to the degenerate surfaces set in the space of all real surfaces of degree 4 in RP^3 .

The Petrovski inequality is still true for nongradient vector fields (see [1]).

It gives the best possible bound for the Poincaré index of a vector field in R^n , whose components are homogeneous polynomials of degree $m - 1$ (A G Hovanski).

6. Maxima singularities

Let

$$F(y) = \max_x f(x, y)$$

be the maxima function $F: B \rightarrow R$ of a family $f: M \times B \rightarrow R$ of real functions on a compact closed manifold M , depending on a parameter y , belonging to a "base space" B , which is an (open) manifold of dimension n .

The maxima function is continuous, but generically, is not smooth. Empirical data lead to a conjecture: the maxima function for a generic family is topologically equivalent (in some neighbourhood of every base space point y) to a Morse function (that is, either to a non-zero linear function or to a sum of a constant and of a non-degenerate quadratic form at point 0).

For $n \leq 6$ this is proved by Briskalova (see [4]).

For the general case, the arguments are:

(i) Suppose for a given $y \in B$ there is only one maximum point x (may be degenerate). In this case the graph of F has at y a tangent plane, and the maxima function is generically topologically linear.

(ii) Suppose for a given $y \in B$ there are $n + 1$ maxima points. Then the graph of F is a pyramid with $n + 1$ faces, each face having at y a tangent plane. The maxima function is generically either topologically linear or topologically equivalent to a Morse function at a minimum point.

(iii) In a generic family the maximum is obtained at s points for y on a manifold of codimension $s - 1$. Along this submanifold S the graph of $F|_S$ has tangent planes and in the transversal direction we can use argument (ii).

(iv) In the particular case of a germ f , which is R^+ -equivalence stable (see [4]) one can prove that the set $\{y, z : z \geq F(y)\}$ is locally diffeomorphic to a convex body. For instance, stability is generic for families with $n \leq 6$ parameters. It is natural to ask whether the set described above is still locally diffeomorphic to a convex body for generical families maxima function singularities, if $n > 6$.

If it is true, this will be one more confirmation for a general principle of fragility of all good things.

To explain this principle let us consider, for example, the set of all polynomials $x^n + a_1x^{n-1} + \dots + a_n$ (a real), having only real zeros (or only non-real, or only zeros with negative real parts). This set, at singularities of its boundary, fills less than one half of the neighbourhood space. Thus under deformation the corresponding property of being good (elliptic, hyperbolic, stable and so on) will rather disappear than persist. The theorems, generated by this principle, (and describing the cones of velocities of curves, moving from the boundary point inside the good set) were recently proved by L. V. Levantovsky.

References

- [1] Arnold V I 1973 Remarks on stationary phase method and Coxeter number; *Usp. Mat. Nauk* 28: 5 17-44
- [2] Arnold V I 1978 Index of vector field singular point, Petrovski-Oleinik inequalities and mixed Hodge structures; *Funkt. Analysis* 12: 1 1-14
- [3] Bernstein I N 1972 Analytical continuation of generalised functions; *Funkt. Analysis* 6: 4 26-40
- [4] Brisgalova L N 1978 On maxima functions of families of functions; *Funkt. Analysis* 12: 1 66-67
- [5] Cerf J 1970 La stratification naturelle des espaces des fonctions differentiables reelles et le theoreme de la pseudo-isotopies; *Institut des Hautes Etudes Sci. Publ. Math.* 39 5-173
- [6] Gousein-Zade S M 1977 Monodromy groups of isolated singularities of hypersurfaces; *Usp. Math. Nauk* 32: 2 23-65
- [7] Hamm H 1977 Genus χ_v for quasi-homogeneous complete intersections; *Funkt. Analysis* 11: 1 87-88
- [8] Harlamov V M 1978 Isotopical classes of nonsingular surfaces of degree 4 in RP^3 ; *Funkt. Analysis* 12: 1 86-87
- [9] Hatcher A E 1973 Parametrised h -cobordism theory; *Ann. Inst. Fourier (Grenoble)* 23: 2 61-74
- [10] Ljashko O V 1976 Decompositions of simple singularities; *Funkt. Analysis* 10: 2 49-56
- [11] Looijenga E 1977 On the semiuniversal deformations of a simple elliptic hypersurface singularity I. *Topology* 16: 3 257-262
- [12] Malgrange B 1974 Integrales asymptotiques et monodromie; *Ann. Sci. Ec. Norm. Suppl. Ser.* 47: 3 405-430

- [13] Pham F 1977 *Caustics and microfunctions* (Kyoto Univ : RIMS) **12** Suppl. pp. 357-366
- [14] Steenbrink J H M 1976 Mixed Hodge structure on vanishing cohomology, *Real and complex singularities*, Oslo
- [15] Steenbrink J H M 1977 Intersections form for quasi-homogeneous singularities; *Compositio Math.* **34** 211-223
- [16] Varchenko A N 1976 Newton polyhedra and estimations of oscillating integrals; *Funkt. Analysis* **10**: 3 13-38
- [17] Volodin I A 1972 Generalised Whitehead groups and pseudo-isotopies; *Usp. Math. Nauk* **27**: 5 229-230
- [18] Wagoner J B 1971 Algebraic invariants for pseudo-isotopies; *Lecture Notes in Mathematics* (Berlin-Heidelberg—New York: Springer) **209** 164-190

Department of Mathematics,
University of Moscow,
Moscow 117 234, USSR.

Yang-Mills and bundles over algebraic curves

By

M F ATIYAH and R BOTT

1. Introduction

The Yang-Mills equations are currently the subject of extensive investigations in theoretical physics in connection with elementary particles. These equations are defined in physical Minkowski space and can be analytically continued to the Euclidean 4-space, where they become elliptic variational equations. Some of the mathematical aspects are treated for example in [1] and [2].

The two-dimensional version of the (Euclidean) Yang-Mills equations are not of direct physical interest but they have considerable mathematical interest and tie in rather unexpectedly with older theories and problems. If we work over the 2-sphere it turns out that Yang-Mills theory becomes essentially identical to the Morse theory for geodesics on compact Lie groups as studied and developed in [4]. If we replace the 2-sphere by a compact orientable surface of genus g we find a new and surprising connection with the theory of stable vector bundles over algebraic curves as developed by Narasimhan and Seshadri [8] and Seshadri [11]. Using an appropriate form of the Morse theory one obtains very precise information on the cohomology of the moduli space of stable bundles. In this way the results of Harder-Narasimhan [6], [7] and Newstead [10] can be re-derived and supplemented.

A comparison between the Morse theory approach and the algebro-geometric method of Harder-Narasimhan is instructive and intriguing. Although the methods are entirely different there is a striking structural parallelism which strongly suggests that the Tamagawa number of a Lie group should have a cohomological interpretation. We shall elaborate on this aspect in § 5.

A full account of this topic is in preparation but it will inevitably be lengthy and so it seemed of interest to present a briefer version at this stage. Detailed proofs will not be given and the treatment will be somewhat discursive, emphasising the main features.

In § 2 we recall basic facts about the Yang-Mills equation and apply them on the 2-sphere. The resulting picture is then compared with the geodesic theory of Lie groups.

In § 3 we proceed to the case of higher genus and discuss the relation with stable bundles. We also give a general formula for computing the Morse indices of the critical sets.

§ 4 is devoted to cohomological results which will be spelled out in detail for the simple case of the group $U(2)$.

Finally in § 5 we compare our methods with those of Harder-Narasimhan which rest on Deligne's proof of the Weil conjectures.

2. Yang-Mills on S^2

The general Yang-Mills set-up is the following. We fix a compact Riemannian manifold M , a compact Lie group G and a principal G -bundle P over M . Then for every connection A in P we have the curvature F , and the Yang-Mills functional or Lagrangian is simply the natural L^2 -norm $\|F\|^2$. This is obtained by integration over M of the local L^2 -norm of F computed using the inner product on G (Killing form) and that on the tangent space to M (defined by the Riemannian metric). The Yang-Mills equations are just the Euler-Lagrange equations of this functional.

When $\dim M = 2$ the curvature F can be viewed as an infinitesimal automorphism of P , i.e. as a section of the vector bundle $\text{ad}(P)$ associated to P by the adjoint representation of G . The Yang-Mills equations assert that this infinitesimal automorphism is *covariant constant*. This implies that the local holonomy group of A lies in the torus $T \subset G$ generated by this infinitesimal automorphism (together with the centre of the Lie algebra of G if G is not semi-simple). In particular when M is the 2-sphere S^2 the global holonomy group lies in T and the Yang-Mills equations for G reduce to those for T . Now for the circle group $U(1)$ the Yang-Mills equations assert that the curvature (which is now just a closed 2-form) is harmonic in the sense of Hodge theory. For a torus $T = U(1) \times \dots \times U(1)$ the Yang-Mills equations assert that each component of the curvature is a harmonic 2-form, and is uniquely determined by the corresponding cohomology class.

To sum up we see that on S^2 we have a complete description of all solutions of the Yang-Mills equations. For $G = U(n)$ we can use the language of vector bundles and say that the solutions are given by direct sums of line-bundles each endowed with its unique "harmonic connection".

For $G = SU(n)$, and more generally for any simply-connected G , the bundle P over S^2 is necessarily trivial topologically. This means that the minimum for the Yang-Mills functional $\|F\|^2$ is given by the zero connection. The other solutions of the Yang-Mills equations correspond to unstable critical points.

If we regard S^2 as the homogeneous space $SU(2)/U(1)$ then it is easy to see that the Yang-Mills connections described above are precisely the *homogeneous connections*, i.e., they are the connections induced from the standard $U(1)$ -connection by a homomorphism $U(1) \rightarrow G$. This shows that the critical points of the Yang-Mills functional over S^2 correspond naturally to the closed geodesics on G (passing through 1). These closed geodesics are themselves the critical points of the "energy" functional $\int |f'(t)|^2 dt$ for maps $f: S^1 \rightarrow G$ (with $f(1) = 1$). This correspondence between the critical points of these two different functionals is not a superficial one. To get a better understanding of this correspondence we must observe that the Yang-Mills functional is defined on the space $\mathcal{A}$ of all G -connections on P and is *invariant* under the group $\mathcal{G}$ of all bundle automorphisms of P (leaving S^2 fixed). Assuming for simplicity that G is simply-connected, $P = S^2 \times G$ and $\mathcal{G} = \text{Map}(S^2, G)$ is the function space of all maps $S^2 \rightarrow G$. The subgroup $\mathcal{G}_0$ consisting of based maps (taking a base point on S^2 to $1 \in G$) acts

freely on the contractible space $\mathcal{A}$ and so $\mathcal{A}/\mathcal{G}_0$ is homotopically the classifying space

$$B\mathcal{G}_0 = B \operatorname{Map}_0(S^2, G) \sim (\operatorname{Map}_0(S^1, G) = \Omega(G).$$

Thus the Yang-Mills functional and the energy functional are defined on spaces of the same homotopy type. The correspondence described above between critical points indicates that the corresponding Morse theories are essentially identical. Now the Morse theory for the energy was studied in [4] and from the homology of the critical manifolds (all of the form $G/C(T)$ where $C(T)$ is the centraliser of a torus T) complete information on the homology of $\Omega(G)$, in particular the absence of torsion, was obtained. In principle, the Yang-Mills functional could be used in precisely the same way and would of course yield the same results. New results will however follow when we replace S^2 by a surface of genus g , although we shall then play the game in reverse by deducing information on the critical manifolds from global information on the function space.

3. Bundles on Riemann surfaces

We now take M to be a compact oriented surface of genus g with a Riemannian metric. If A is a G -connection which satisfies the Yang-Mills equations then as in § 2 the curvature generates a torus T in G , the local holonomy lies in T while the global holonomy lies in the centraliser $C(T)$ of T . Thus we have a homomorphism ρ of the fundamental group $\pi_1(M)$ into $C(T)/T$. Except for problems to do with (finite) group extensions we see that the connection A breaks up into a torus part, which has to be harmonic as in § 2, and a flat part given by the homomorphism $\rho : \pi_1(M) \rightarrow C(T)/T$. For example when $G = U(2)$ there are just two cases:

(i) $\dim T = 1$, T is the centre of $U(2)$ and

$$C(T)/T = SU(2)/\{\pm 1\} \cong SO(3),$$

(ii) $\dim T = 2$, T is the maximal torus and $C(T) = T$.

In the first case all the information is carried by the homomorphism $\rho : \pi_1(M) \rightarrow SO(3)$ and we have two further alternatives depending on whether ρ lifts to $SU(2)$ or not. In case (ii), on the other hand, ρ is trivial and all the information is contained in the two line-bundles.

We see therefore that flat bundles occur naturally amongst the Yang-Mills solutions and that Morse theory is likely therefore to give information on the spaces of moduli of such flat bundles. The interest of this lies in the relation between flat bundles and stable algebraic bundles over the algebraic curve defined by the conformal structure of M . We recall the main result of [8] according to which an algebraic vector bundle of rank n , and Chern class 0, over M is stable, in the sense of Mumford, if and only if it arises from an irreducible unitary representation of $\pi_1(M)$. Moreover isomorphic bundles correspond to equivalent representations. For bundles with non-zero Chern class there are corresponding results involving projective unitary representations. Thus the moduli spaces of stable bundles, which are of algebro-geometric interest, can be identified topologically with moduli

spaces of representations. Hence use of Morse theory applied to the Yang-Mills functional for $U(n)$ will yield information about the cohomology of the moduli spaces of stable vector bundles.

Further light is cast on this whole question if we observe that *any* unitary G -connection defines a holomorphic G° -bundle (G° being the complexification of G). We use the connection to define an obvious almost complex structure on the principal G° -bundle and then appeal to the Newlander-Nirenberg integrability theorem [9] to see that we have a holomorphic structure (for more details see [2]). Conversely the holomorphic structure of the G° -bundle together with a reduction to G leads uniquely to the original G -connection. In particular, when $G = U(n)$ this means that a unitary connection is the same as a holomorphic vector bundle endowed with a hermitian metric. To reach a Yang-Mills connection we can now try to fix the holomorphic structure and vary the hermitian metric (or the reduction to G in the general case) so as to obtain a constrained minimum for $\|F\|^2$. When $M = S^2$ the theorem of Grothendieck [5] asserts that we have a discrete classification given by a reduction to the maximal torus. The constrained minimum in each class is unique and does in fact yield a Yang-Mills solution. Moreover all solutions arise in this way. Thus the classification of Yang-Mills solutions coincides with Grothendieck's holomorphic classification which in turn coincides with the homogeneous classification.

For surfaces of genus g the situation is somewhat more complicated for two reasons. In the first place the holomorphic classification is not discrete, but has moduli. Secondly not every constrained minimum leads to a Yang-Mills connection: one still has to further minimise with respect to the appropriate moduli. Nevertheless the general picture is that, for "most" connections, the process of keeping the holomorphic structure fixed and varying the metric will lead eventually to a Yang-Mills minimum ($F \equiv 0$). This explains why one should expect the general or stable holomorphic bundle to arise from a representation of the fundamental group.

For a Yang-Mills connection A one can introduce the Morse index, namely the number of negative eigenvalues in the second variation of the Yang-Mills functional. This index is always finite and is given by a simple formula which we now describe. Recall that the curvature F of A can be viewed as a section of $\text{ad}(P)$. The Lie bracket with iF is a self-adjoint endomorphism of $\text{ad}(P)$ which is a covariant constant. Thus the subspace $\text{ad}^+(P)$ of $\text{ad}(P) \otimes \mathbb{C}$ on which this endomorphism is positive is a holomorphic sub-bundle. The formula for the index is then:

$$(3.1) \quad \text{index } A = 2c_1(\text{ad}^+P) + 2 \dim(\text{ad}^+P)(g - 1)$$

where c_1 denotes the Chern class.

This formula arises by first identifying the negative space of the second variation with the sheaf cohomology group $H^1(M, \text{ad}^-P)$, where ad^-P is defined analogously to ad^+P (and in fact $\text{ad}^-P \cong (\text{ad}^+P)^*$). Combined with the Riemann-Roch theorem and the vanishing in this case of H^0 we end up with the formula (3.1).

As an example consider $G = U(2)$ and A corresponding to a direct sum of line-bundles L_1, L_2 each with its harmonic connection.

Assume $c_1(L_1) = k_1, c_1(L_2) = k_2$ with $k_1 > k_2$.

Then $\text{ad}^+(P) \cong L_1 \otimes L_2^*$,

so that $\text{index } A = 2\{(k_1 - k_2) + (g - 1)\}$.

Since $k_1 + k_2 = c_1(P)$ is fixed we can rewrite this as

$$(3.2) \quad \text{index } A = 4k_1 - 2c_1(P) + 2g - 2.$$

In the next section we shall use this formula for the index in our cohomological calculations. It is important to note that (3.1) is a topological formula in that it is unchanged under continuous variation of A (in the space of Yang-Mills solutions).

4. Cohomological computations

As a preliminary example, we consider the Morse theory for $SU(2)$ -bundles over S^2 . The Yang-Mills connections are then given by the direct sum $L^* \oplus L^{-*}$ where $k = 0, 1, 2, \dots$ and L is the standard Hopf bundle with its standard harmonic connection. In the function space $\mathcal{A}/\mathcal{G}_0$ each value of $k \geq 1$ contributes a 2-sphere, this being the orbit of $G = SU(2)$ acting on a given critical point. For $k = 0$ on the other hand we have a single point representing the trivial connection, which is the absolute minimum. For $k \geq 1$ the Morse index, as given by formula (3.2), is $4k - 2$. We can now combine all the information on the homology of the critical sets by introducing the Morse series $\mathcal{M}_t$ in which each critical manifold X contributes $t^v P_t(X)$ where v is the Morse index and P_t is the Poincaré polynomial. In our case we get therefore

$$(4.1) \quad \mathcal{M}_t = 1 + \sum_{k=1}^{\infty} t^{4k-2} (1 + t^2) = 1/(1 - t^2),$$

which coincides with the Poincaré series of the function space $\mathcal{A}/\mathcal{G}_0 \sim \Omega(SU(2))$. This is precisely what would be predicted by Morse theory if one could show analytically that it was applicable in this function space situation. In fact for the corresponding energy functional the validity of the Morse theory is classical and so it is reasonable to expect it to hold for the 2-dimensional Yang-Mills. In any case we shall assume that this is so, not only for S^2 but for all surfaces, and examine the cohomological consequences.

Note next that if we replace $SU(2)$ by $U(2)$ we have one component for each integer l , corresponding to the first Chern class. The case $l = 0$ is essentially that of $SU(2)$ while for $l = 1$ all critical manifolds are 2-spheres and the index of the k th sphere, as given by (3.2), is now $4k$, giving

$$(4.2) \quad \mathcal{M}_t = \sum_{k=0}^{\infty} t^{4k} (1 + t^2) = 1/(1 - t^2).$$

This is consistent with the fact that all components of $\Omega(U(2))$ have the same homotopy type and therefore the same Poincaré series.

Before proceeding to the case of higher genus we must digress to discuss the effect of symmetry in Morse theory. If a functional is invariant under a group acting freely then we naturally pass to the quotient by the group action, as we have done in factoring $\mathcal{A}$ by $\mathcal{G}_0$. If the action is not free, as with the whole action of $\mathcal{G}$ on $\mathcal{A}$ or equivalently with the action of $G = \mathcal{G}/\mathcal{G}_0$ on $\mathcal{A}/\mathcal{G}_0$, we can take account of the group action in cohomological computations by using equivariant

cohomology. This is essentially equivalent to considering the Morse theory of the induced function on bundles with fibre as the original manifold and the symmetry group as the structure group. For example if we have in $\mathcal{A}/\mathcal{G}_0$ a critical manifold which is a G -orbit, with isotropy group H , then its contribution to the equivariant version of the Morse series is defined as $t^v P_t(B_H)$ where v is the index. In some cases, e.g., $G = SU(2)$ and $H = S^1$ we have

$$(4.3) \quad P_t(B_H) = P_t(G/H) \cdot P_t(B_G),$$

in which case the equivariant Morse theory we are describing reduces to the ordinary Morse theory. This is the situation with Yang-Mills on S^2 but on surfaces of higher genus we will encounter critical manifolds which are acted on freely by G (i.e., with $H = 1$) and (4.3) is then false. Thus in such cases the good formulae are the equivariant ones.

After these remarks we consider the case when M has genus g , $G = U(2)$ and the first Chern class is equal to 1. This is the simplest interesting case. The Yang-Mills connections are then of two types, namely

(i) the absolute minima where the curvature is in the centre of $U(2)$ and is harmonic,

(ii) sums of harmonic line-bundles $L_1 \oplus L_2$.

The connections of type (ii) fall into different components depending on the Chern classes k_1, k_2 of L_1 and L_2 ($k_1 + k_2 = 1$ by hypothesis). Each component in $\mathcal{A}/\mathcal{G}_0$ is a product $S^2 \times J \times J$ where J is the Jacobian of the Riemann surface, and $G = U(2)$ acts on the first factor. The isotropy group is therefore $U(1) \times U(1)$. The Morse index is given by (3.2) and so the contribution to the equivariant Morse series is

$$t^{4k_1+2g-4} \frac{(1+t)^{4g}}{(1-t^2)^2} \quad (\text{where } k_1 > k_2 = 1 - k_1).$$

Summing over k_1 we see that the total contribution from all critical points of type (ii) is

$$(4.4) \quad \frac{t^{2g} (1+t)^{4g}}{(1-t^2)^2 (1-t^4)}.$$

The space of absolute minima has only the central $U(1)$ of $U(2)$ as isotropy group and therefore contributes

$$(4.5) \quad \frac{1}{1-t^2} P_t(R(g)),$$

to the equivariant Morse series. Here we have denoted by $R(g)$ the space of all connections (up to isomorphism) of type (i). Thus the equivariant Morse series is given by

$$(4.6) \quad \mathcal{M}_t = \frac{1}{1-t^2} \left\{ P_t(R(g)) + \frac{t^{2g} (1+t)^{4g}}{(1-t^2)(1-t^4)} \right\}.$$

To compute the equivariant Poincaré series of the function space, namely, $P_t(B\mathcal{G})$, we express the Riemann surface M homotopically in the form

$$M \sim S^1 \vee \dots \vee S^1 \vee e^2,$$

where there are $2g$ copies of S^1 and e^2 is a 2-cell attached appropriately. This gives rise to a fibration sequence for $\mathcal{G}$ and hence for $B\mathcal{G}$ from which one eventually concludes that

$$(4.7) \quad P_t(B\mathcal{G}) = \frac{\{(1+t)(1+t^3)\}^{2g}}{(1-t^2)^2(1-t^4)}.$$

If we now assume that we are in the fortunate situation in which $\mathcal{M}_t = P_t$, then we can equate (4.6) and (4.7) to derive

$$(4.8) \quad \frac{\{(1+t)(1+t^3)\}^{2g}}{(1-t^2)(1-t^4)} = P_t(R(g)) + \frac{t^{2g}(1+t)^{4g}}{(1-t^2)(1-t^4)},$$

which yields a formula for the Betti numbers of the manifold $R(g)$. To justify this assumption we need the analytical aspects of the Morse theory (as before) and, in addition, some topological argument to show that there can be no cancellation between terms in the Morse series, which *a priori* only dominates the Poincaré series. Since odd powers of t are now present (unlike the case of S^2) such cancellation is theoretically possible. However one can show that all the putative homology arising from critical manifolds can in fact be "completed" so that cancellation does not occur. Subject therefore to our basic analytical hypothesis about the validity of the Morse theory we have found in (4.8) a formula for the Betti numbers of $R(g)$. It is already interesting and not self-evident that (4.8) does indeed lead to a *polynomial* (of degree $8g-6$) with non-negative coefficients. For example,

$$g = 1, P_t(R(g)) = (1+t)^2,$$

$$g = 2, P_t(R(g)) = (1+t)^4 \{1+t^2+4t^3+t^4+t^6\}.$$

The interest of the manifold $R(g)$ is that it can be identified with the moduli space of algebraic rank 2 stable vector bundles over M with Chern class 1. This is quite easy to show directly and is very close to the Narasimhan-Seshadri approach [9] to stable bundles. The presence of the factor $(1+t)^{2g}$ in $P_t(R(g))$ arises because we have not so far normalised the determinant line bundle. In other words taking the determinant gives a fibre map $R(g) \rightarrow J$ in which the fibre $S(g)$ represents stable bundles with fixed determinant. This is the moduli space of primary interest and formulae for its Betti numbers are given in [6] and [7]. To arrive directly at the homology of $S(g)$ by our method we have to apply the Morse theory to the restricted space of $U(2)$ -connections in which the central component is a fixed harmonic connection (of Chern class 1). The result is that $P_t(S(g))$ is given by formula (4.8) with a factor $(1+t)^{2g}$ removed. In other words

$$(4.9) \quad P_t(R(g)) = P_t(S(g)) \cdot P_t(J).$$

This formula is equivalent to the homological triviality of the fibration $R(g) \rightarrow J$. Now using the tensor product action of J on $R(g)$, one finds that $R(g)$ is the flat bundle over J associated to the action of $\pi_1(J) = \mathbb{Z}^{2g}$ on $S(g)$ via the homomorphism $\pi_1(J) \rightarrow J_2 = \mathbb{Z}_2^{2g}$, where $J_2 \subset J$ is the group of elements of period 2. Thus (4.9) is equivalent to

$$(4.10) \quad J_2 \text{ acts trivially on the cohomology of } S(g).$$

This is in fact a special case of the main theorem of [7] where it is proved by comparing the theory for $U(2)$ and $PU(2)$. Our argument here is formally similar. On the other hand the geometrical meaning of (4.10) is quite clear in our context because the homology of both $S(g)$ and $R(g)$ appears as parts of the homology of the corresponding function space. The action of $\pi_1(J)$ on the homology of $S(g)$ is then part of its action on the homology of the function space, and this is easily seen to be trivial.

For larger Lie groups such as $U(n)$ for $n \geq 3$ we can proceed in a similar fashion. The Yang-Mills minimum gives the relevant moduli space of stable bundles while the other critical points are expressible in terms of line bundles and Yang-Mills minima of smaller groups $U(m)$ with $m < n$. By an induction process one can in principle obtain formulae generalising (4.8) but the details are complicated and will not be pursued here. One difficulty should however be mentioned and that concerns bundles with 0 Chern class. The moduli space of stable bundles in this case is well known not to be compact [8], and its natural compactification tends to be singular. Even if we tried to avoid this case in studying $U(3)$ -bundles we would inevitably meet it amongst the sub-ordinate $U(2)$ -bundles, corresponding to other critical points.

As mentioned earlier our Morse theory approach not only yields the Betti numbers of $S(g)$, it also shows that these spaces have no torsion. The logical structure of the argument is interesting because it essentially consists of relating the Morse theory for general genus to that of genus 0. More precisely information about the critical manifolds in the S^2 case leads to information about the function space $\Omega(G)$. This is then fed into the Morse theory for the general surface M and we retrace the argument to derive finally information about the critical manifolds for M .

5. Comparison with Harder-Narasimhan

As mentioned earlier our Morse Theory approach to the cohomology of moduli spaces of stable bundles is quite different but, in many ways, strikingly parallel to the method of Harder and Narasimhan ([6], [7]). In order to bring out this parallelism we shall first need to review briefly the components of the Harder-Narasimhan method.

They start from the Weil conjectures, established in their final form by Deligne, which relate Betti numbers of a complex algebraic variety to the numbers of points of the "corresponding" variety over finite fields. In particular, the Betti numbers of the moduli spaces of stable bundles over curves are related to the number of stable bundles over a curve defined over a finite field. On the other hand, the function field analogue of the classical Siegel formula (cf. [12]), yields a formula for a weighted sum over all bundles, stable and unstable. For $U(2)$ the unstable bundles can be described in terms of line bundles and so their contribution can be explicitly summed. Substituting this into the Siegel formula then gives the number of stable bundles. For $U(n)$ with $n \geq 3$ the unstable bundles have to be treated inductively and the results are more complicated.

The Siegel formula involves the use of adelic groups G_A for the function field K of a curve Y over a finite field F_q . For rank n bundles $G = GL_n$ and the

bundles on Y , defined over F_q , are in bijective correspondence with the double coset space $\mathcal{K} \backslash G_A / G_K$ where $\mathcal{K}$ is a maximal compact subgroup of G_A . If we fix the determinant of the vector bundle then GL_n is replaced by SL_n and there is a different $\mathcal{K}$ for each of the different determinant line bundles. One now introduces the Tamagawa measure τ on G_A and on its quotient G_A / G_K (recall that G_K is discrete in G_A). The total measure of G_A / G_K is called the Tamagawa number of G (over K) and for SL_n this is known to be 1. The Siegel formula is now obtained by computing the measure of G_A / G_K using its decomposition into $\mathcal{K}$ -orbits, i.e., into the double cosets $\mathcal{K} \backslash G_A / G_K$. We get

$$(5.1) \quad 1 = \sum_a \tau(\mathcal{K} / \mathcal{K}_a),$$

where a runs over the double cosets and $\mathcal{K}_a$ is the isotropy group, which is *finite* because it is the intersection of the discrete group G_K with a conjugate of the compact group $\mathcal{K}$. We can rewrite (5.1) in the form

$$(5.2) \quad 1/\tau(\mathcal{K}) = \sum_a 1/|K_a|.$$

Interpreted in terms of vector bundles with given determinant $\mathcal{K}_a$ is the automorphism group of the corresponding bundle. For a stable bundle this reduces to the multiplicative scalars F_q^* . The value of $\tau(\mathcal{K})$ is known in terms of the ζ -function of Y ; for SL_n we have

$$(5.3) \quad \tau(\mathcal{K}) = q^{-(n^2-1)(g-1)} \zeta(2)^{-1} \dots \zeta(n)^{-1}$$

where $g = \text{genus}(Y)$, and $\zeta(s)$ has the usual form

$$(5.4) \quad \zeta(s) = \frac{\prod_{i=1}^{2g} (1 - \omega_i q^{-s})}{(1-q)^{-s} (1-q \cdot q^{-s})}$$

where $|\omega_i| = q^{1/2}$.

Putting these formulae together, and decomposing the sum in (5.2) into the stable and unstable part, we obtain a formula for the number of stable bundles. Considering this formula as a function of q (i.e., replacing F_q by F_{q^n}) and using the Weil conjectures leads finally to the Betti number formula for the moduli space of stable bundles.

The parallel with our method is now fairly clear. In both cases we start with an infinite-dimensional space (G_A / G_K and the space $\mathcal{A}$ of connections) on which an infinite-dimensional group acts ($\mathcal{K}$ and $\mathcal{G}$ respectively). We then compute Tamagawa measure and equivariant cohomology respectively. In both cases the initial space is, in the appropriate sense, trivial namely

$$(5.5) \quad \tau(G_A / G_K) = 1,$$

$$P_t(\mathcal{A}) = 1,$$

where P_t is the ordinary Poincaré series (recall that $\mathcal{A}$ is contractible). The results of both calculations involve a sum over stable and unstable bundles, and the individual parts of the formulae look alike in the two situations.

All of this suggests that the Weil conjectures, which relate homology to numbers of rational points, have some extension to the case of *infinite-dimensional* spaces,

with Tamagawa measure replacing number. Particularly noteworthy is the analogy between the groups $\mathcal{K}$ and $\mathcal{Q}$. The formulae (5.3) and (5.4) are very close to those of $P_+(B\mathcal{Q})$ (the case of $U(2)$ is given in (4.7) and $U(n)$ is similar). Since $\mathcal{K}$ is the product over all places p of Y of the groups G_{K_p} (over the completions K_p) the analogy with $\mathcal{Q}$ is quite clear.

It would be very interesting to understand this analogy in more detail. In particular examination of (5.5) shows that the formula for the Tamagawa number should have some cohomological significance. There are a number of interesting ideas and speculations in this direction, but a fuller understanding remains as a challenging problem.

References

- [1] Atiyah M F, Drinfeld V G, Hitchin N J and Manin Yu I 1978 Construction of instantons. *Phys. Lett.* **A65** 185–187
- [2] Atiyah M F, Hitchin N J and Singer I M 1978 Self-duality in four-dimensional Riemannian geometry. *Proc. R. Soc. (London)* **A362** 425–461
- [3] Atiyah M F and Jones J D S 1978 Topological aspects of Yang-Mills theory. *Commun. Math. Phys.* **61** 97–118
- [4] Bott R 1956 An application of the Morse theory to the topology of Lie groups. *Bull. Soc. Math. France* **84** 251–281
- [5] Grothendieck A 1957 Sur la classification des fibres holomorphes sur la sphere de Riemann. *Am. J. Math.* **79** 121–138
- [6] Harder G 1970 Eine Bemerkung Zu einer Arbeit von P E Newstead. *J. Math.* **242** 16–25
- [7] Harder G and Narasimhan M S 1975 On the cohomology groups of moduli spaces of vector bundles over curves. *Math. Ann.* **212** 215–248
- [8] Narasimhan M S and Seshadri C S 1965 Stable and unitary vector bundles on a compact Riemann surface. *Ann. Math.* **82** 540–567
- [9] Newlander A and Nirenberg L 1957 Complex analytic co-ordinates in almost complex manifolds. *Ann. Math.* **65** 391–404
- [10] Newstead P E 1967 Topological properties of some spaces of stable bundles. *Topology* **6** 241–262
- [11] Seshadri C S 1967 Space of unitary vector bundles on a compact Riemann surface. *Ann. Math.* **85** 303–336
- [12] Weil A 1961 Adeles and algebraic groups. Lecture Notes, Princeton

M F ATIYAH,
Mathematical Institute,
24–29, St. Giles,
Oxford OX 1 3LB, UK

and

R BOTT,
Department of Mathematical Science,
Harvard University,
Cambridge, Mass 02138, USA.

Vanishing theorems for square-integrable harmonic forms

By

JOZEF DODZIUK*

There are two ways of proving vanishing theorems for cohomology of compact Riemannian manifolds. One is based on Bochner's lemma, a variant of maximum principle, the other is an application of Weitzenböck identity and integration by parts. Both are discussed in great detail in [9]. In this paper we show that the latter method yields vanishing theorems for L^2 -harmonic forms on complete manifolds. The extra ingredient required in the proof is due to Andreotti and Vesentini [1]; (see also [4], theorem 26).

Some of the results of this paper have been obtained in a stronger form in [8] using a generalisation of maximum principle; however our proofs seem to be more straightforward. Corollary 1 has also been obtained by Yau [10]. Our aim is to show that machinery developed to study harmonic forms on compact manifolds works equally well for complete ones. We do obtain some new results regarding L^2 -harmonic spinors and a generalisation of Matsushima's vanishing theorems [7].

The author is very grateful to Professor E. Calabi for numerous discussions on the subject, in particular for suggesting that Matsushima's theorem might generalise to complete manifolds and for explaining it to the author in terms of Riemannian geometry.

Throughout this paper M will denote a complete, oriented Riemannian manifold of n dimensions. We will denote by $\langle \cdot, \cdot \rangle$ and $|\cdot|$ the inner product and the norm induced in fibers of various tensor bundles by the metric of M . $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ will denote the global (integrated) inner product and norm respectively. We will use the standard notation for the Laplacian Δ , exterior derivative d and volume element dV .

At this point we recall the Weitzenböck identity ([4], p. 131) for the Laplacian acting on a p -form α , $p = 0, 1, 2, \dots, n$.

$$(1) \quad \nabla \alpha = \nabla^* \nabla \alpha + F_p \alpha,$$

where ∇ denotes the operator of covariant differentiation, ∇^* is its formal adjoint and F_p is an algebraic operator involving the curvature tensor of M linearly (see [4], p. 131 for explicit formulae). We now state two theorems whose proof will be given simultaneously.

*Partially supported by National Science Foundation Grant No. MPS 75-07917.

Theorem 1. Suppose the form $A_p(a) = \langle F_p a, a \rangle$ is positive semi-definite at every point of M . Then every L^2 -harmonic p -form is parallel. If either the volume of M is infinite or there exists a point $x \in M$ such that A_p is strictly positive at x , then every L^2 -harmonic form is identically zero.

Theorem 2. Suppose the form A_p is bounded from below at every point $x \in M$, i.e., $A_p(a) \geq -C|a|^2$ for some constant $C > 0$ independent of x and every $a \in \Lambda^p T_x^* M$. Let ω be a C^∞ p -form on M such that both ω and $\Delta\omega$ are square-integrable. Then $\nabla\omega$ is in L^2 , the functions $\langle \nabla^* \nabla \omega, \omega \rangle$, and $A_p(\omega) = \langle F_p \omega, \omega \rangle$ are in L^1 , and we have

$$(\nabla^* \nabla \omega, \omega) = (\nabla \omega, \nabla \omega)$$

$$(\Delta \omega, \omega) = (\nabla \omega, \nabla \omega) + (F_p \omega, \omega).$$

Proof (cf. [4], theorem 26). In order to carry out the familiar proof given when M is compact one has to justify integration by parts, i.e., the use of Stokes' theorem. To do this one constructs a family of cut-off functions whose supports exhaust M .

Let $\mu : (0, \infty) \rightarrow \mathbf{R}$ be a smooth function such that $0 \leq \mu \leq 1$, $\text{supp } \mu \subset [0, 2]$ and $\mu(t) = 1$ for $t \in [0, 1]$. Let $d(x)$ be the geodesic distance between $x \in M$ and a fixed point $x_0 \in M$. For every $R > 0$ set

$$\lambda_R(x) = \mu\left(\frac{d(x)}{R}\right).$$

We denote by B_R the geodesic ball of radius R centred at x_0 . It is easy to see that for every $R > 0$, λ_R is Lipschitz; hence differentiable almost everywhere on M , and has the following properties.

$$0 \leq \lambda_R(x) \leq 1 \quad \text{for every } x \in M$$

$$\text{supp } \lambda_R \subset B_{2R}$$

$$(2) \quad \lambda_R(x) = 1 \quad \text{for every } x \in B_R$$

$$\lim_{R \rightarrow \infty} \lambda_R = 1$$

$$|d\lambda_R(x)| \leq C_1/R \quad \text{almost everywhere on } M,$$

where $C_1 > 0$ is a constant independent of R . We also observe that because M is complete $\bar{B}_R$ is compact for every $R > 0$.

We now compute $(\Delta\omega, \lambda_R^2 \omega)$ for a C^∞ square-integrable p -form ω . By (1) and Stokes' theorem, which is applicable to Lipschitz continuous forms, we have

$$(3) \quad (\Delta\omega, \lambda_R^2 \omega) = (\lambda_R \nabla \omega, \lambda_R \nabla \omega) + (\lambda_R \nabla \omega, 2d\lambda_R \otimes \omega) + (F_p \omega, \lambda_R^2 \omega).$$

Moreover, from (2) and the inequality $|(a, b)| \leq \frac{1}{2}(\|a\|^2 + \|b\|^2)$ we obtain

$$(4) \quad \|d\lambda_R \otimes \omega\|^2 = \int_M |d\lambda_R \otimes \omega|^2 dV \leq \int_M |d\lambda_R|^2 |\omega|^2 dV \leq \frac{C_1^2}{R^2} \|\omega\|^2,$$

and

$$(5) \quad |(\lambda_R \nabla \omega, 2d\lambda_R \otimes \omega)| \leq \frac{1}{2} \|\lambda_R \nabla \omega\|^2 + \frac{2C_1^2}{R^2} \|\omega\|^2.$$

Theorem 1 now follows easily. If $\Delta\omega = 0$, (3) and (5) imply that

$$\frac{1}{2} \|\lambda_R \nabla \omega\|^2 + (F_p \lambda_R \omega, \lambda_R \omega) \leq \frac{2C_1^2}{R^2} \|\omega\|^2.$$

Notice that both summands on the left are non-negative. Taking the limit as R tends to infinity we see that $\|\nabla \omega\| = 0$, i.e. ω is parallel. Thus $|\omega|$ is constant and has to be zero if the volume is infinite. Finally

$$(F_p \omega, \omega) = \int_M \langle F_p \omega, \omega \rangle dV$$

exists and is equal to zero. It follows that $\omega(x) = 0$ for every point x , where the form A_p is strictly positive. If such a point exists ω is identically zero.

Theorem 2 is proved similarly. From (3) and (5), assuming that $\|\omega\| < \infty$, $\|\Delta\omega\| < \infty$, we obtain

$$\begin{aligned} \|\lambda_R \nabla \omega\|^2 &= -(\lambda_R \nabla \omega, 2d\lambda_R \otimes \omega) - (F_p \lambda_R \omega, \lambda_R \omega) + (\Delta\omega, \lambda_R^2 \omega) \\ &\leq \frac{1}{2} \|\lambda_R \nabla \omega\|^2 + \frac{2C_1^2}{R^2} \|\omega\|^2 - (F_p \lambda_R \omega, \lambda_R \omega) + (\Delta\omega, \lambda_R^2 \omega). \end{aligned}$$

Further, since $\langle F_p \alpha, \alpha \rangle \geq -C|\alpha|^2$,

$$\frac{1}{2} \|\lambda_R \nabla \omega\|^2 \leq \frac{2C_1^2}{R^2} \|\omega\|^2 + C \|\lambda_R \omega\|^2 + (\Delta\omega, \lambda_R^2 \omega),$$

which proves that $\|\nabla \omega\| < \infty$ by taking the limit as R approaches infinity. Now by (1) at every point $x \in M$ we have

$$(6) \quad \langle \nabla^* \nabla \omega, \omega \rangle = \langle \Delta\omega, \omega \rangle - \langle F_p \omega, \omega \rangle.$$

According to assumptions of theorem 2

$$-A_p(\omega) = -\langle F_p \omega, \omega \rangle \leq C|\omega|^2. \text{ Therefore}$$

$$|\langle \nabla^* \nabla \omega, \omega \rangle| \leq |\langle \Delta\omega, \omega \rangle| + C|\omega|^2.$$

Since both terms on the right are in L^1 so is $\langle \nabla^* \nabla \omega, \omega \rangle$, and in view of (6) $\langle F_p \omega, \omega \rangle$ is also integrable. To complete the proof of theorem 2 we consider $(\nabla^* \nabla \omega, \lambda_R \omega)$. Using Stokes' theorem again we see that

$$(\nabla^* \nabla \omega, \lambda_R \omega) = (\nabla \omega, \lambda_R \nabla \omega) + (\nabla \omega, d\lambda_R \otimes \omega).$$

Passing to the limit as R tends to infinity and using (4), Lebesgue-dominated convergence theorem and the fact that $\nabla \omega$ is in L^2 we conclude that $(\nabla^* \nabla \omega, \omega) = (\nabla \omega, \nabla \omega)$ which finishes the proof.

Remarks. (a) A very similar reasoning shows that for every C^∞ form ω on M with $\|\omega\| < \infty$, $\|\Delta\omega\| < \infty$ we have

$$(\Delta\omega, \omega) = (d\omega, d\omega) + (\delta\omega, \delta\omega).$$

In particular every L^2 -harmonic form is closed and co-closed. As a matter of fact, our proof of both theorems above is patterned after the proof of this result given in [4].

(b) Theorems analogous to the above hold for L^2 -holomorphic forms on complete-Kählerian manifolds and for L^2 -harmonic spinors on complete Riemannian manifolds with a spin structure. As regards harmonic spinors, there are two possible ways of defining them: either as solutions of equation $D\alpha = 0$, where D is the Dirac operator, or as spinors annihilated by the spinor-Laplacian D^*D . The two definitions are equivalent for L^2 -spinors, as the identity

$$(D^*D\alpha, \alpha) = (D\alpha, D\alpha)$$

shows. This identity, which holds for C^∞ spinors α with $\|\alpha\| < \infty$, $\|D^*D\alpha\| < \infty$ can be proved easily by an argument analogous to the one employed in the proof of theorems 1 and 2.

As in the compact case one obtains specific vanishing theorems by identifying the algebraic operator F_p in special cases. We give a few examples.

Corollary 1 (cf. [9], [10]). Let M be a complete, oriented, Riemannian manifold with positive semi-definite Ricci curvature. Every L^2 -harmonic form of degree one is parallel. Moreover, there are no non-zero L^2 -harmonic 1-forms if either the volume of M is infinite or the Ricci curvature is positive definite at some point of M .

Corollary 2 (cf. [9]). On a complete Kählerian manifold of positive semi-definite Ricci curvature every L^2 -holomorphic form is parallel. There are no L^2 -holomorphic forms if volume of M is infinite. If the Ricci curvature is strictly positive at some point of M , every L^2 -holomorphic form of positive degree is identically zero.

Corollary 3 (cf. [5], [6]). Let M be a complete spin manifold of non-negative scalar curvature. Then every L^2 -harmonic spinor on M is parallel. There are no L^2 -harmonic spinors if either the volume of M is infinite or if the scalar curvature is positive at some point of M .

We now consider a generalisation of vanishing theorems of Matsushima [7] to non-compact manifolds. We begin by describing an unpublished interpretation of Matsushima's theorem in terms of Riemannian geometry due to Calabi (cf. [2]). We fix $0 < p < \dim M$. The Riemannian metric induces covariant differentiation ∇ in various tensor bundles on M . In particular,

$$\nabla \cdot \nabla : C^\infty(\Lambda^p T^*M) \rightarrow C^\infty(T^*M \otimes T^*M \otimes \Lambda^p T^*M).$$

We denote by $\nabla \wedge \nabla$ the composition of $\nabla \cdot \nabla$ with the bundle map

$$T^*M \otimes T^*M \otimes \Lambda^p T^*M \rightarrow \Lambda^2 T^*M \otimes \Lambda^p T^*M$$

arising from alternating of the first two factors. By Ricci identities $\nabla \wedge \nabla \omega = K\omega$ is an algebraic operator depending linearly on curvature tensor of M . More precisely, in terms of classical tensor notation (cf. [4], p. 130),

$$(K\omega)_{a,b,i_1,\dots,i_p} = (\nabla_a \nabla_b - \nabla_b \nabla_a) \omega_{i_1,\dots,i_p} = \sum_{k=1}^p R_{ik}^j \omega_{i_1,\dots,i_{k-1}^j i_{k+1},\dots,i_p},$$

where the summation over repeated indices is understood and R_{ik}^j are components of the curvature tensor. Assuming M is compact and ω is a C^∞ p -form, we can proceed as follows.

$$\|K\omega\|^2 = (K\omega, \nabla \wedge \nabla \omega) = (K\omega, \nabla \cdot \nabla \omega),$$

since at every point $K\omega$ is orthogonal to all tensors $\xi_{ab i_1,\dots,i_p}$ which are symmetric in a, b and alternating in $i_1, i_2, \dots, i_p$. Hence

$$\|K\omega\|^2 = (\nabla^* K\omega, \nabla \omega).$$

Now an easy computation yields

$$\begin{aligned} (\nabla^* K\omega)_{b i_1,\dots,i_p} &= -\frac{1}{2} \sum_k (\nabla^a R_{ab} j_{i_k}) \omega_{i_1,\dots,i_{k-1}^j i_{k+1},\dots,i_p} \\ &\quad - \frac{1}{2} \sum_k R_{ik}^j \nabla^a \omega_{i_1,\dots,i_{k-1}^j i_{k+1},\dots,i_p}. \end{aligned}$$

At this point one makes the assumption that the curvature tensor of M satisfies $\nabla^a R_{abij} \equiv 0$, which is certainly true if M is locally symmetric, i.e., in the case considered by Matsushima. Thus we can write

$$(7) \quad \nabla^* K\omega = Q \nabla \omega,$$

where Q is the endomorphism of the bundle $T^*M \otimes \Lambda^p T^*M$ given by

$$(Q\xi)_{b i_1,\dots,i_p} = -\frac{1}{2} \sum_k R_{ik}^j \xi_{a i_1,\dots,i_{k-1}^j i_{k+1},\dots,i_p}.$$

Since M is assumed to be compact there exists a constant $\beta > 0$ such that $|Q\xi| \leq \beta |\xi|$ for every $x \in M$ and every $\xi \in (T^*M \otimes \Lambda^p T^*M)_x$.

In this setting Matsushima's theorem takes the following form.

Theorem (Calabi). Suppose M is a compact, oriented Riemannian manifold with curvature tensor satisfying $\nabla^i R_{ijkl} = 0$. Let $\beta > 0$ be a constant such that $Q(\xi, \xi) \leq \beta |\xi|^2$ for every $x \in M$ and every $\xi \in (T^*M \otimes \Lambda^p T^*M)_x$. If the quadratic form

$$\Phi_p(\omega) = \frac{1}{\beta} \langle K\omega, K\omega \rangle + \langle F_p \omega, \omega \rangle,$$

is positive definite at every point $x \in M$, then every harmonic p -form on M is identically zero.

Proof: Suppose $\Delta\omega = 0$. Then, using (1), Stokes' theorem and boundedness of Q , we have

$$\begin{aligned} 0 &= \|\nabla \omega\|^2 + \langle F_p \omega, \omega \rangle \geq \frac{1}{\beta} (Q \nabla \omega, \nabla \omega) + \langle F_p \omega, \omega \rangle \\ &= \int_M \Phi_p(\omega) dV. \end{aligned}$$

Since Φ_p is assumed to be positive definite everywhere ω is identically zero.

A generalisation of the above to open manifolds is

Theorem 3. Let M be a complete, oriented Riemannian manifold satisfying the following conditions.

- (a) The curvature tensor of M satisfies $\nabla^i R_{ijkl} = 0$.
- (b) There exists $C > 0$ such that $\langle F_p a, a \rangle \geq -C |a|^2$ for every $x \in M$, $a \in T_x^* M$.
- (c) There exists $\beta > 0$ such that $Q(a, a) \leq \beta |a|^2$ for every $x \in M$ and every $a \in (T^* M \otimes \Lambda^2 T^* M)_x$.
- (d) The quadratic form Φ_p is positive definite at every point $x \in M$.

Under these conditions every L^2 -harmonic form on M is identically 0.

Remark : (a) (b), (c) are automatically satisfied if M is locally symmetric.

Proof : As before we have to justify the integration by parts. Suppose ω is an L^2 -harmonic form. Because of (b) theorem 2 applies and

$$0 = (\omega, \Delta \omega) = (\nabla \omega, \nabla \omega) + (F_p \omega, \omega).$$

Applying (c) we obtain

$$0 \geq \frac{1}{\beta} (Q \nabla \omega, \nabla \omega) + (F_p \omega, \omega).$$

The theorem will be proved if we show that $(Q \nabla \omega, \nabla \omega) = (K \omega, K \omega)$. We consider $(\lambda_R K \omega, \lambda_R K \omega)$, where $\{\lambda_R\}_{R>0}$ is the family of cut-off functions used in the proof of theorems 1 and 2. Applying Schwartz inequality, Stokes' theorem and (7), we have

$$\begin{aligned} (8) \quad \|\lambda_R^2 K \omega\|^2 &= (\lambda_R^2 K \omega, \nabla \wedge \nabla \omega) = (\lambda_R^2 K \omega, \nabla \cdot \nabla \omega) = \\ &= (\nabla^* K \lambda_R^2 \omega, \nabla \omega) = (Q \nabla (\lambda_R^2 \omega), \nabla \omega) = \\ &= (Q (2\lambda_R d\lambda_R \otimes \omega), \nabla \omega) + (\lambda_R^2 Q \nabla \omega, \nabla \omega). \end{aligned}$$

Using (c) and (4) we can estimate the first summand on the right as follows.

$$\begin{aligned} |(Q (2\lambda_R d\lambda_R \otimes \omega), \nabla \omega)| &\leq \beta \|2\lambda_R d\lambda_R \otimes \omega\| \cdot \|\nabla \omega\| \\ &\leq 2\beta \|d\lambda_R \otimes \omega\| \cdot \|\nabla \omega\| \leq 2\beta \frac{C_1}{R} \|\omega\| \cdot \|\nabla \omega\|. \end{aligned}$$

This implies that $(K \omega, K \omega) = (Q \nabla \omega, \nabla \omega)$ and together with (b) proves the theorem.

Remark : A similar theorem holds for L^2 -holomorphic forms on complete Kählerian manifolds. We will denote appropriate quadratic forms by Ψ_p .

Calabi (unpublished, cf. [2]) studied quadratic forms Ψ_p on bounded symmetric domains D in $\mathbb{C}^n$. According to his calculation $\Psi_p > 0$ if $0 < p < \gamma(D)$, where

$\gamma(D)$ is given below (see [3] for classification of bounded symmetric domains and notation).

D	$\gamma(D)$
$I_{m,m'}$	$\min(m, m')$
$I_m (m \geq 4)$	$m - 1$
III_m	$(m + 1)/2$
IV_m	$m/2$
V	6
VI	9

In view of theorem 3 this yields

Corollary 4. Let M be a Kählerian manifold covered by a bounded symmetric domain D in $\mathbb{C}^n$. If $0 \leq p < \gamma(D)$, then M admits no non-zero L^2 -holomorphic p -forms.

References

- [1] Andreotti A and Vesentini E 1965 Carleman estimates for the Laplace-Beltrami equation on complex manifolds. *Publ. Math. IHES* **25** 1-30
- [2] Calabi E 1963 Matsushima's theorem in Riemannian and Kählerian geometry. *Notices Am. Math. Soc.* **10** 505
- [3] Calabi E and Vesentini E 1960 On compact locally symmetric Kähler manifolds. *Ann. Math.* **71** 472-507
- [4] de Rham G 1973 *Variétés différentiables* (Paris : Hermann)
- [5] Hitchin N 1974 Harmonic spinors. *Adv. Math.* **14** 31-55
- [6] Lichnerowicz A 1973 Spineurs harmoniques. *CR Acad. Sci. Paris Ser AB* **257** 7-9
- [7] Matsushima Y 1962 On the first Betti number of compact quotients of higher-dimensional symmetric spaces. *Ann. Math.* **75** 312-330
- [8] Vesentini E 1967 Lectures on Levi convexity. Tata Institute of Fundamental Research, Bombay
- [9] Yano K and Bochner S 1953 Curvature and Betti numbers. *Ann. Math. Studies* No. 32, (Princeton : University Press)
- [10] Yau S T 1976 Some function theoretic properties of complete Riemannian manifolds and their application to geometry. *Indiana Univ. Math. J.* **25** 659-670.

Department of Mathematics,
University of Pennsylvania,
Philadelphia,
Pennsylvania 19174, USA

On operators of trace class in $L^2(X, \mu)$

By

J J DUISTERMAAT

It is well-known that if X is an n -dimensional C^∞ manifold, μ a positive C^∞ density on X and A an integral operator on X with compactly supported C^∞ distribution kernel K on $X \times X$, then A is of trace class as an operator in $L^2(X, \mu)$ and $\text{Tr } A = \int_X K(x, x) d\mu$. For the definition and basic facts about Hilbert-Schmidt operators and operators of trace class, see Gelfand and Wilenkin ([1], Kap. 1, §2).

On the other hand, if X is an n -dimensional topological manifold, μ a positive continuous density on X , then there exist integral operators A with compactly supported continuous kernel K which are not of trace class in $L^2(X, \mu)$. To see this, let U be a non-void coordinate neighbourhood in X , which can be identified with an open subset of $T = \mathbb{R}^n/\mathbb{Z}^n$. Because multiplication by $\mu^{1/2}$ is an isometry: $L^2(U, \mu) \rightarrow L^2(U, 1)$ which does not change the class of operators with continuous kernels having compact support $\subset U \times U$, we may assume that $\mu = 1$. If every such integral operator is of trace class then using translations and a partition of unity in T it would follow that every integral operator on T with continuous kernel is of trace class. Now let A be convolution with the continuous function a on T . If A is of trace class then the following series is absolutely convergent:

$$\begin{aligned} & \sum_{\xi \in \mathbb{Z}^n} (A \exp(2\pi i \langle \cdot, \xi \rangle), \exp(2\pi i \langle \cdot, \xi \rangle)), \\ &= \sum_{\xi \in \mathbb{Z}^n} \iint a(x-y) \exp(2\pi i \langle x-y, \xi \rangle) dy dx \\ &= \sum_{\xi \in \mathbb{Z}^n} \int a(z) \exp(2\pi i \langle z, \xi \rangle) dz, \end{aligned}$$

which is just the Fourier series of a . However, Du Bois-Reymond already in 1876 constructed $a \in C(\mathbb{R}/\mathbb{Z})$ with divergent Fourier series, see Zygmund ([3], Ch. 8, § 1), and this example easily lifts to $\mathbb{R}^n/\mathbb{Z}^n$.

Nevertheless the feeling is that an integral operator with continuous kernel is very close to being of trace class, and the theorem stated below supports this feeling.

Let X be a topological space, $BC(X)$ the space of bounded continuous functions: $X \rightarrow \mathbb{C}$ with the supremum norm. Assume that μ is a positive Radon measure on X with finite total measure, that is $\mu \in BC(X)'$, $\mu(f) > 0$ if $f \geq 0$, $f \neq 0$.

Then $L^2(X, \mu)$ is the completion of $BC(X)$ with respect to the hermitian inner product $(f, g) = \mu(f \cdot \bar{g})$, we assume that $L^2(X, \mu)$ is separable. The formula $(j(g))(f) = \mu(f \cdot g)$ defines an isometry $j: L^2(X, \mu) \rightarrow L^2(X, \mu)'$, dual spaces of normed linear spaces will be provided with the strong topology defined by the operator norm.

Noting that an integral operator with bounded continuous kernel is continuous: $BC(X)' \rightarrow BC(X)$, the following theorem shows that the question whether such operators are of trace class is related to the question what Hilbert spaces can be continuously embedded into $BC(X)$.

Theorem. Let H be a Hilbert space, i a continuous linear mapping: $H \rightarrow BC(X)$. Then, if A is a continuous linear mapping: $H' \rightarrow H$, the operator $\tilde{A} = i \circ A \circ i' \circ j$ is of trace class as an operator in $L^2(X, \mu)$, it is an integral operator with bounded and separately continuous kernel K , the function $x \mapsto K(x, x)$ is Lebesgue-integrable with respect to μ and finally

$$\text{Tr } \tilde{A} = \int_X K(x, x) d\mu.$$

Proof. Consider i as a continuous linear mapping: $H \rightarrow L^2(X, \mu)$. Then $E = (i \circ i^*)^{1/2}$ is a continuous positive self-adjoint operator in $L^2(X, \mu)$ with range equal to $i(H) \subset BC(X)$, and $E^{-1} \circ i$ induces an isometry between $H/\text{Ker } i$ and the Hilbert subspace $\overline{i(H)}$ of $L^2(X, \mu)$. In fact, E is uniquely determined by these properties (see Pijls [2], theorem 1.3.2).

Because E is continuous: $L^2(X, \mu) \rightarrow BC(X)$ it is an integral operator with kernel $\mathcal{E}$ such that $x \mapsto \mathcal{E}(x, \cdot) = j^{-1} \circ E'(\delta_x)$ is a bounded function on X with values in $L^2(X, \mu)$. Here δ_x denotes the Dirac measure at $x \in X$. If e_k is an orthonormal basis in $L^2(X, \mu)$ then

$$(1) \quad j^{-1} \circ E' \delta_x = \sum_k (j^{-1} \circ E' \delta_x, e_k) e_k = \sum_k \delta_x(\overline{Ee_k}) \cdot e_k,$$

with convergence in $L^2(X, \mu)$, the partial sums being uniformly bounded in $L^2(X, \mu)$ with respect to $x \in X$. Hence

$$(2) \quad (\mathcal{E}(y, \cdot), \mathcal{E}(x, \cdot)) = \sum_k (Ee_k)(x) \cdot (\mathcal{E}(y, \cdot), e_k) \\ = \sum_k (Ee_k)(x) \cdot (\overline{Ee_k})(y),$$

the partial sums being uniformly bounded in x and y . In particular the partial sums of

$$x \mapsto (\mathcal{E}(x, \cdot), \mathcal{E}(x, \cdot)) = \sum_k (Ee_k)(x) \cdot (\overline{Ee_k})(x)$$

are a bounded family in $BC(X)$, by the Lebesgue-dominated convergence theorem the pointwise limit is Lebesgue integrable with respect to μ . But this just means that $\mathcal{E} \in L^2(X \times X, \mu \times \mu)$, that is E is a Hilbert-Schmidt operator. Because $B = E^{-1} \circ i \circ A \circ (E^{-1} \circ i)'$ is continuous: $L^2(X, \mu)' \rightarrow L^2(X, \mu)$, it follows that $\tilde{A} = E \circ B \circ E' \circ j$ is of trace class in $L^2(X, \mu)$.

In view of the compactness of E , the e_k can be chosen as eigenvectors of E for eigenvalues $\lambda_k \geq 0$. Continuing (1) we find

$$(3) \quad j^{-1} E' \delta_{\theta} = \sum_k \lambda_k \cdot \overline{\delta_{\theta}(e_k)} \cdot e_k = \sum_k \overline{e_k(x)} \cdot E e_k.$$

Applying $E \circ B \circ j$ and using $j \circ E = E' \circ j$ gives

$$(4) \quad K(y, x) = (E \circ B \circ E'(\delta_{\theta}))(y) = \sum_k (\tilde{A} e_k)(y) \cdot \overline{e_k(x)},$$

converging uniformly in y and the partial sums being uniformly bounded in x . Again applying the Lebesgue dominated-convergence theorem it follows $x \mapsto K(x, x) \in L^1(X, \mu)$ and is approximated in the $L^1(X, \mu)$ -sense by the partial sums of

$$x \mapsto \sum_k (\tilde{A} e_k)(x) \cdot \overline{e_k(x)},$$

which is a bounded family in $BC(X)$. Therefore

$$(5) \quad \int_X K(x, x) d\mu = \sum_k (\tilde{A} e_k, e_k) = \text{Tr } \tilde{A},$$

and the proof is complete.

Remarks. An easy consequence is the theorem of Mercer: if A is a positive self-adjoint operator in $L^2(X, \mu)$, and at the same time an integral operator with bounded continuous kernel K , then it is of trace class and $\text{Tr } A = \int_X K(x, x) d\mu$. The proof is by taking $H = \text{range of } A^{1/2}$, which maps $BC(X)'$ into $L^2(X, \mu)$ and $L^2(X, \mu)$ into $BC(X)$ (both continuously), and providing H with the inner product

$$(f, g)_H = (A^{-1} f, A^{-1} g), \quad A^{-1} : H \rightarrow (\text{Ker } A)^{\perp}.$$

If X is an n -dimensional C^{∞} manifold, then for any compact subset U of X the Sobolev space $H_{\text{loc}}^{s/2}(X) \cap \mathcal{E}'(U)$ has the topology of a Hilbert space and is contained in $C_{\text{comp}}(X)$ with continuous inclusion if (and only if) $s > n$. (The subscripts "loc" resp. "comp" mean that the functions locally belong to the given space, resp. have compact support and belong to the given space.) Because the dual space of $H_{\text{comp}}^{s/2}(X)$ is equal to $H_{\text{loc}}^{-s/2}(X)$, the theorem can be applied to any continuous linear mapping

$$A : H_{\text{loc}}^{-s/2}(X) \rightarrow H_{\text{comp}}^{s/2}(X), \quad s > n.$$

This in turn is the case if A has distribution kernel

$$K \in H_{\text{comp}}^s(X \times X), \quad s > n,$$

thus giving back the opening statement of the article. The operator E in the proof is then a pseudo-differential operator of order $-s/2$. If one is not fussy about the order s one may work in the framework of linear partial differential operators and take for E the parametrix of a sufficiently high integral power of the Laplace operator on X with respect to some Riemannian metric.

It may be finally noted that expansions like (4) can be used to prove that $C_{\text{comp}}^{\infty}(Y \times X)$ is equal to the topological tensor product of $C_{\text{comp}}^{\infty}(X)$ and $C_{\text{comp}}^{\infty}(Y)$, for arbitrary finite-dimensional C^{∞} manifolds X, Y , this in turn leads to the Schwartz kernel theorem.

References

- [1] Gelfand I M and Wilenkin N J 1964 *Verallgemeinerte Funktionen IV* (Berlin : VEB)
- [2] Pijls H G J 1975 *Locally convex algebras in the spectral theory and eigenfunction expansions*, Thesis, Amsterdam
- [3] Zygmund A 1968 *Trigonometric series* (Cambridge : University Press)

Mathematisch Instituut der
Rijksuniversiteit,
Budapestlaan,
De Uithof, Utrecht,
The Netherlands

Deformations of metrics and associated harmonic maps

By

J EELLS and L LEMAIRE

1. Introduction

In this paper we establish a theorem ((3.1) below) concerning the dependence of harmonic maps on the Riemannian metrics used to define them. Our method is an application of the implicit function theorem in suitable manifolds of maps.

In the case of a compact range of negative curvature, that result was obtained by Sampson [24], using derivative estimates for harmonic maps given in [10]. Another special case has been found by Schoen-Yau [26], and applied by them to obtain theorem (6.3) below. Their proof is entirely different; it is based on a Dirichlet growth theorem of Morrey ([21] lemma 9.4.18) and thus is restricted to maps with two-dimensional domains.

In some cases the geometry of the problem introduces irrelevant degeneracy. For instance, that can happen if M or N has a nontrivial group of isometries—or if the harmonic map under consideration factorises through a geodesic. More technical results concerning some of these cases are given in § 5.

As applications, we offer proofs of two theorems on harmonic diffeomorphisms of surfaces, due to Sampson [24] and Schoen-Yau [26].

We are much indebted to Professor Sampson for many stimulating conversations, as well as for making available to us (many years ago) a preliminary version of his manuscript [24]—a primary source of motivation for the present work.

2. Basic properties

(2.1) Let M and N be smooth (i.e. C^∞) finite dimensional manifolds, with M compact. Unless otherwise specified, we suppose M and N without boundary. We denote by $T(M)$ and $T^*(M)$ the tangent and cotangent vector bundles of M .

We are primarily interested in smooth Riemannian metrics on M and N and their associated smooth harmonic maps. However, our main tool is the implicit

During the preparation of this paper the authors enjoyed the hospitality of the Institut des Hautes Etudes Scientifiques. The first-named thanks the Université de Paris VI, where he was Professeur d'Echange. The second-named is Chargé de recherches au Fonds National Belge de la Recherche Scientifique.

function theorem, which in its simplest form applies to maps between Banach spaces. Therefore, at times we shall restrict ourselves to finite orders of differentiability. Furthermore, we need to solve certain linear elliptic equations, and we choose to do so in Schauder's context; see for instance [12]. Therefore, for r an integer such that $2 \leq r \leq \infty$ and $0 < \alpha < 1$, we shall consider the space $C^{r+\alpha}(M, N)$ of maps whose differentials of orders $\leq r$ satisfy an α -Hölder condition. With the associated topology, $C^{r+\alpha}(M, N)$ is a smooth manifold modelled on a Banach space, provided $r < \infty$.

Let $M^{r+\alpha}$ and $N^{r+\alpha}$ be the spaces of $C^{r+\alpha}$ -Riemannian metrics on M and N . These are open subsets of the Fréchet spaces $C^{r+\alpha}(\odot^2 T^*(M))$ and $C^{r+\alpha}(\odot^2 T^*(N))$. If $r < \infty$, $M^{r+\alpha}$ is an open subset of a Banach space, since M is compact.

(2.2) Given $(g, h) \in M^{r+1+\alpha} \times N^{r+1+\alpha}$ and a $C^{r+2+\alpha}$ -map $\varphi: M \rightarrow N$ (sometimes written $\varphi: (M, g) \rightarrow (N, h)$), its *energy* is defined by

$$E(\varphi) = E(\varphi; g, h) = \frac{1}{2} \int_M |d\varphi(x)|^2 dx.$$

Then $E: C^{r+2+\alpha}(M, N) \rightarrow \mathbb{R} (\geq 0)$ is a $C^{r+1+\alpha}$ -function.

(2.3) The extremals of E are called *harmonic maps*, and satisfy the Euler-Lagrange equation [10]

$$\tau_\varphi = 0; \text{ i.e.}$$

$$\nabla_\bullet E(\varphi) = - \int_M \langle \tau_\varphi(x), v(x) \rangle dx = 0$$

for all variations $v \in T_\varphi(C^{r+2+\alpha}(M, N))$ of φ . τ_φ is an element of $T_\varphi(C^{r+\alpha}(M, N))$ called the *tension field* of φ . At times we shall have to specify the metrics used and write $\tau_\varphi = \tau(\varphi; g, h)$. In local coordinate systems $\{x^i\}$ and $\{u^\alpha\}$ on M and N , τ_φ is represented by

$$\tau^\gamma_\varphi(\varphi) = g^{ij} \left(\frac{\partial^2 \varphi^\gamma}{\partial x^i \partial x^j} - {}^M \Gamma^k_{ij} \frac{\partial \varphi^\gamma}{\partial x^k} + {}^N \Gamma^\gamma_{\alpha\beta}(\varphi) \frac{\partial \varphi^\alpha}{\partial x^i} \frac{\partial \varphi^\beta}{\partial x^j} \right),$$

where ${}^M \Gamma$ and ${}^N \Gamma$ are the Christoffel symbols of g and h .

Given a harmonic map $\varphi: M \rightarrow N$, the second variation of its energy is described by the *Hessian* $\nabla^2 E(\varphi)$, characterised by the identity [20], [29]

$$\nabla^2 E(\varphi)(v, w) = \int_M \langle J_\varphi v(x), w(x) \rangle dx,$$

where $-J_\varphi$ is the *Jacobi operator* along φ , represented in the notations of ([9], §3) by

$$-J_\varphi(v) = \text{Trace } \nabla_\varphi^2 v + \text{Trace } R^N(d\varphi, v) d\varphi,$$

where R^N is the curvature tensor of (N, h) . J_φ is a linear elliptic operator. Integration by parts, using the symmetry properties of R^N , shows that

$$\int_M \langle J_\varphi v, w \rangle dx = \int_M \langle v, J_\varphi w \rangle dx; \text{ i.e.,}$$

that J_φ is self-adjoint.

The condition that $\nabla^2 E(\varphi)$ is non-degenerate is that $\text{Ker } J_\varphi = 0$ (or equivalently $\text{Ker } J_\varphi^* = 0$). The first equality implies that φ is an isolated critical point of E and the second that

$$J_\varphi : T_\varphi(\mathbf{C}^{r+2+\alpha}(M, N)) \rightarrow T_\varphi(\mathbf{C}^{r+\alpha}(M, N))$$

surjective. Indeed, the equation $J_\varphi v = V$ can be solved iff $V \perp \text{Ker } J_\varphi^*$. The on-degeneracy of $\nabla^2 E(\varphi)$ implies therefore that J_φ is an isomorphism.

(2.4) For our application of the implicit function theorem, we consider the tension field τ defined in (2.3) as a map

$$\tau : \mathbf{C}^{r+2+\alpha}(M, N) \times \mathbf{M}^{r+1+\alpha} \times \mathbf{N}^{r+k+1+\alpha} \rightarrow T(\mathbf{C}^{r+\alpha}(M, N)).$$

We work in a neighbourhood of a C^∞ -harmonic map $\varphi_0 : (M, g_0) \rightarrow (N, h_0)$ where g_0 and h_0 are C^∞ , but in order to remain in the context of Banach spaces, we take $r < \infty$ and restrict attention to an exponential chart $\mathbf{U}$ of $\mathbf{C}^{r+2+\alpha}(M, N)$ centred at φ_0 and such that the image of every map in the chart is contained in a fixed compact domain K of N . By the canonical identification through $\exp_{\varphi_0}$ (using the metric h_0), we have

$$T(\mathbf{C}^{r+\alpha}(M, N)) \Big|_{\mathbf{U}} = \mathbf{U} \times T_{\varphi_0}(\mathbf{C}^{r+\alpha}(M, N)).$$

We can represent τ by its composition with the projection on the second factor, which we still denote by τ :

$$(2.5) \quad \tau : \mathbf{U} \times \mathbf{M}^{r+1+\alpha} \times \mathbf{N}^{r+k+1+\alpha} \rightarrow T_{\varphi_0}(\mathbf{C}^{r+\alpha}(M, N)).$$

τ is now a C^k -map from an open set of a Banach space to a Banach space.

(2.6) *Lemma. Consider a smooth path*

$$(\varphi_s; g_s, h_s)_{s \in I} \text{ in } \mathbf{U} \times \mathbf{M}^{r+1+\alpha} \times \mathbf{N}^{r+k+1+\alpha} \Big|_K$$

starting at the given C^∞ -triple $(\varphi_0; g_0, h_0)$ and put

$$v = \frac{\partial \varphi_s}{\partial s} \Big|_{s=0}, \quad X = \frac{\partial g_s}{\partial s} \Big|_{s=0}, \\ Y = \frac{\partial {}^M \Gamma_s}{\partial s} \Big|_{s=0} \quad \text{and} \quad Z = \frac{\partial {}^N \Gamma_s}{\partial s} \Big|_{s=0},$$

where ${}^M \Gamma_s$ and ${}^N \Gamma_s$ are the Christoffel symbols associated to g_s and h_s . (Y depends on X , and is given by

$$Y_{ij}^k = \frac{1}{2} g^{kp} (\nabla_i X_{jp} + \nabla_j X_{ip} - \nabla_p X_{ij}).$$

If $\varphi_0 : (M, g_0) \rightarrow (N, h_0)$ is harmonic, then

$$(2.7) \quad \frac{d\tau}{ds} \Big|_{s=0} = d\tau(\varphi_0; g_0, h_0)(v, X, Z) = \Delta v + \text{Trace } R^N(d\varphi_0, v) d\varphi_0 +$$

$$\text{Trace}_X(\nabla^2 \varphi_0) - \text{Trace}(Y d\varphi_0) + \text{Trace } Z(d\varphi_0, d\varphi_0).$$

The terms depending on v are derived in [29], and the others in [19].

3. The main dependence theorem

(3.1) *Theorem* : Suppose that $\varphi_0 : (M, g_0) \rightarrow (N, h_0)$ is a C^∞ -harmonic map between C^∞ -Riemannian manifolds such that $\nabla^2 E(\varphi_0)$ is non-degenerate. Then for $1 \leq k < \infty$ and $r < \infty$ there is a neighbourhood V of (g_0, h_0) in $M^{r+1+\alpha} \times N^{r+k+1+\alpha}$ and a unique C^k -map $S : V \rightarrow C^{r+2+\alpha}(M, N)$ such that $S(g_0, h_0) = \varphi_0$ and $S(g, h)$ is a harmonic map $\varphi : (M, g) \rightarrow (N, h)$.

From (2.7) we note that the hypothesis on the Hessian is satisfied if the sectional curvature Riem^N is strictly negative and $\varphi_0(M)$ is not a point or a geodesic.

(3.2) *Corollary* : Let $k \geq 1$ and $r \leq \infty$. Consider a C^k -path $(g_s, h_s)_{s \in I}$ of C^∞ -metrics in $M^{r+1+\alpha} \times N^{r+k+1+\alpha}$ and suppose that for each $s \in I$ there is a unique harmonic map $\varphi_s : (M, g_s) \rightarrow (N, h_s)$ in a fixed homotopy class, with each $\nabla^2 E(\varphi_s)$ non-degenerate. Then the map $I \rightarrow C^{r+2+\alpha}(M, N)$ given by $s \rightarrow \varphi_s$ is C^k .

If $r = \infty$, we interpret (3.2) as saying that the map is continuous into $C^\infty(M, N)$ and induces a C^k -map into $C^{r+2+\alpha}(M, N)$.

(3.3) *Remark* : In particular, the conclusion of (3.2) is valid if

- (a) each h_s has negative sectional curvature;
- (b) each (N, h_s) satisfies the growth conditions described in ([9]; (5.2.a))—always satisfied if N is compact;
- (c) the homotopy class contains neither constant maps nor maps to closed geodesics.

Application : Such continuous dependence in dimension 2 has been used by Earle-Eells (6) to construct a specific continuous section of the fibre bundle of complex structures over the Teichmüller space of a closed surface.

Proof of theorem (3.1).

Consider the map τ defined as in (2.5). From (2.7) and the non-degeneracy of $\nabla^2 E(\varphi_0)$, we deduce that its partial differential in the U direction is an isomorphism.

By the implicit function theorem in Banach spaces [5] there is a neighbourhood V of (g_0, h_0) in $M^{r+1+\alpha} \times N^{r+k+1+\alpha}|_K$ and a unique C^k -map $S : V \rightarrow C^{r+2+\alpha}(M, N)$ such that $S(g_0, h_0) = \varphi_0$ and $\tau(S(g, h); g, h) = 0$ for all $(g, h) \in V$.

The corollary follows easily by covering the path with neighbourhoods V constructed as in the theorem and using the unicity of the harmonic maps φ_s .

When $r = \infty$ the assertion follows from the characterisation of the C^∞ -topology on $C^\infty(M, N)$ as the smallest one making each inclusion map $C^\infty(M, N) \rightarrow C^k(M, N)$ continuous.

(3.4) *Remark*. Real analytic ($= C^\omega$) harmonic maps have properties not generally shared by other C^ω -maps [32]. Let M and N be C^ω -manifolds. It is known ([31], [23]) that the C^ω -metrics are dense in $M^\infty \times N^\infty$. Furthermore, if $\varphi : (M, g) \rightarrow (N, h)$ is a harmonic map between C^ω -Riemannian manifolds, then $\varphi \in C^\omega$. Theorem (3.1) shows that any C^∞ -harmonic map φ between C^ω -manifolds

with C^∞ -Riemannian metrics for which $\nabla^2 E(\varphi)$ is non-degenerate can be C^∞ -approximated by C^ω -maps which are harmonic with respect to suitable metrics.

4. Dependence for the Dirichlet problem

(4.1) Suppose now that M has a boundary ∂M and consider the Dirichlet problem

(4.2) $\tau\varphi = 0$

$$\varphi|_{\partial M} = \psi$$

for a fixed map $\psi \in E^\infty(\partial M, N)$, the space of restrictions to ∂M of maps in $C^\infty(M, N)$. We denote by $C_\psi^{r+\alpha}(M, N)$ the Banach manifold of maps $\varphi : M \rightarrow N$ coinciding with ψ on ∂M , $r < \infty$. Its tangent bundle

$$T(C_\psi^{r+\alpha}(M, N)) = \{v \in T(C^{r+\alpha}(M, N)) : v|_{\partial M} = 0 \text{ and } \pi \circ v|_{\partial M} = \psi\},$$

where $\pi : T(N) \rightarrow N$ is the canonical bundle projection.

The Jacobi operator $J\varphi$ is now a map

$$T\varphi(C_\psi^{r+2+\alpha}(M, N)) \rightarrow T\varphi(C^{r+\alpha}(M, N)).$$

For $v, w \in T\varphi(C_\psi^{r+2+\alpha}(M, N))$,

integration by parts as in § 2 shows that

$$\int_M \langle J\varphi v, w \rangle dx = \int_M \langle v, J\varphi w \rangle dx,$$

and the non-degeneracy of $\nabla^2 E(\varphi)$ means again that $J\varphi$ is an isomorphism.

(4.3) With these modifications in the proofs, *theorem (3.1) and corollary (3.2) remain valid when the harmonic maps are solutions of the Dirichlet problem (4.2), replacing*

$$C^{r+2+\alpha}(M, N) \text{ by } C_\psi^{r+2+\alpha}(M, N).$$

Furthermore, remark (3.3) remains valid for the Dirichlet problem without conditions (b) and (c). Indeed, the existence and unicity of the solution to that problem are guaranteed simply by the curvature assumption (a) ([25]; the unicity follows also from [15]).

(4.4) In [13] Goldstein establishes the smooth dependence on ψ of the solution of (4.2) in a neighbourhood of a harmonic map, under the hypothesis that $\text{Riem}^N \leq 0$. His method uses the inverse function theorem and applies to the case of a harmonic map with non-degenerate Hessian (without hypothesis on the curvature). Adapting his proof, we see that in a neighbourhood in $C^{r+2+\alpha}(M, N)$ of a map φ_0 with $\nabla^2 E(\varphi_0)$ non-degenerate, the map

$$\Lambda : M^{r+1+\alpha} \times N^{r+k+1+\alpha} \times C^{r+2+\alpha}(M, N) \rightarrow$$

$$M^{r+1+\alpha} \times N^{r+k+1+\alpha} \times T\varphi_0 C^{r+\alpha}(M, N) \times E^{r+2+\alpha}(\partial M, N)$$

defined by $\Lambda(g, h, \varphi) = (g, h, \text{proj. } \tau(\varphi; g, h), \varphi|_{\partial M})$ is a diffeomorphism; here $\text{proj. } \tau(\varphi; g, h)$ denotes the projection defined in (2.4). It follows that if

$\varphi_0 (M, g_0) \rightarrow (N, h_0)$ is a C^∞ -solution of (4.2) with boundary data ψ_0 such that $\nabla^2 E(\varphi_0)$ is non-degenerate, then there is a neighbourhood V of $(g_0, h_0; \psi_0)$ in $M^{r+1+\alpha} \times N^{r+k+1+\alpha} \times C^{r+2+\alpha}(M, N)$ and a C^k -map $V \rightarrow C^{r+2+\alpha}(M, N)$ such that $S(g_0, h_0; \psi_0) = \varphi_0$; and $S(g, h, \psi)$ is a solution of (4.2) for all $(g, h; \psi) \in V$. See ([16]; Thm 10.5.3) for an alternative treatment in the case $N = \mathbf{R}$.

(4.5) *Example.* A simple instance of degeneracy of $\nabla^2 E(\varphi)$ is the following. Take $M = I$, $N = S^n$ (= the Euclidean n -sphere, $n > 1$) and ψ_0 carrying the endpoints of I onto two antipodal points of S^n . In this case the solutions of (4.2) are geodesic segments φ . We observe that φ is not a continuous function of ψ at ψ_0 .

(4.6) *Problem.* When is $\{\psi \in E^\infty(\partial M, N) : E : C_\psi^\infty(M, N) \rightarrow \mathbf{R} \text{ is non-degenerate at every critical point}\}$ residual (i.e., a countable intersection of open dense subsets) in the space of ψ for which (4.2) has a solution? See ([22], ch. VII) for the case $M = I$.

(4.7) *Problem.* In the same order of ideas, when is $\{(g, h) \in M^\infty \times N^\infty : \text{the function } \varphi \rightarrow E(\varphi; g, h) \text{ is non-degenerate}\}$ residual? In that problem one may have to discard certain degeneracies, as was done by Abraham in [2].

(4.8) *Example:* Let $\varphi : M \rightarrow N$ be a harmonic map. A theorem of Visik [30] applied to the operator J_φ asserts that there is a number $\epsilon > 0$ such that for any smooth domain $M' \subset M$ of volume $M' < \epsilon$ we have $\nabla^2 E(\varphi|_{M'})$ non-degenerate.

5. Killing degeneracies

(5.1) Degeneracy of $\nabla^2 E(\varphi; g, h)$ is expressed by the presence of non-zero Jacobi fields along φ . Amongst those are the Killing fields of (M, g) and (N, h) which induce non-zero variations of φ . Let $I(M, g)$ denote the identity component of the isometry group of (M, g) and $\mathbf{I}(M, g)$ its Lie algebra of Killing fields. For $X \in \mathbf{I}(M, g)$ the induced variation $u \in T_\varphi(C^\infty(M, N))$ is defined by $u(x) = d\varphi(x)X(x)$ for all $x \in M$. For $Y \in \mathbf{I}(N, h)$ it is given by $v(x) = Y(\varphi(x))$.

(5.2) There is a substantial difference in those two cases—partly due to the following phenomenon: if S, M, N are smooth manifolds with S and M compact, then the composition map

$$C^r(S, M) \times C^{r+k}(M, N) \rightarrow C^r(S, N),$$

defined by $(\varphi, \psi) \rightarrow \psi \circ \varphi$ is C^k , and not more. [1], [5, Ch. VIII, § 12], [8, § 6.c].

We shall now show that we can remove some degeneracy in these two cases.

(5.3) *Proposition.* Let $\varphi_0 : (M, g_0) \rightarrow (N, h_0)$ be a harmonic map between C^∞ -Riemannian manifolds in a homotopy class consisting of surjective maps. Suppose that $\text{Ker } J_{\varphi_0} = \mathbf{I}(N, h_0)$. Then there exists a neighbourhood V of g_0 in $M^{r+1+\alpha}$ and a C^∞ -map $S : V \rightarrow C^{r+2+\alpha}(M, N)/\mathbf{I}(N, h_0)$ such that $S(g)$ is a family of harmonic maps $\varphi : (M, g) \rightarrow (N, h_0)$.

Proof. As in (5.2), $\mathbf{I}(N, h_0)$ acts smoothly on $C^{r+2+\alpha}(M, N)$. By the hypothesis of surjectivity, this action is free on the component of φ_0 in that space.

For a chart U centred at φ_0 , let $\tilde{U} = U/I(N, h_0)$ denote the orbit manifold of $I(N, h_0)$. Because the tension τ is invariant under the action of $I(N, h_0)$, it induces a C^∞ -map

$$\tilde{\tau} : \tilde{U} \times \mathbf{M}^{r+1+\alpha} \rightarrow T_{\tilde{\varphi}_0} (C^{r+\alpha}(M, N)/I(N, h_0))$$

as in (2.5). Since $\text{Ker } J_{\varphi_0} = \mathbf{I}(N, h_0)$, the partial differential of $\tilde{\tau}$ at $\tilde{\varphi}_0$ in the $\tilde{U}$ direction is an isomorphism. Therefore, there is a neighbourhood V of g_0 in $\mathbf{M}^{r+1+\alpha}$ and a unique C^∞ -map $S : V \rightarrow \tilde{U}$ such that $S(g_0) = \tilde{\varphi}_0$ and $\tilde{\tau}(S(g), g) = 0$ for all $g \in V$. Following $S(g)$ by the action of $I(N, h_0)$, we obtain the S of the proposition.

(5.4) By way of contrast, we see from (5.2) that $I(M, g_0)$ acts continuously but not differentiably on $C^{r+2+\alpha}(M, N)$. Consequently, we are unable to proceed as in (5.3). However, a satisfactory result can be obtained in an important case, found in the classical literature under the heading of Poincaré's continuation method (see [22; Ch. IX] and [27; § 21]).

The case under consideration is that of a non-trivial closed geodesic represented as a harmonic map $\varphi : S^1 \rightarrow (N, h)$. Now the identity component $I(S^1) = S^1$ induces a one-dimensional degeneracy of $\nabla^2 E(\varphi)$, but does not act differentiably on $C^{r+2+\alpha}(S^1, N)$. We follow the usual practice in calling φ non-degenerate if its only Jacobi fields are those of $I(S^1)$.

The following result is surely known (We express our thanks to W Klingenberg and D Sullivan for their comments.):

(5.5) *Proposition. The non-degenerate closed geodesics of (N, h) depend smoothly on small deformations of h .*

Indeed, we identify the prime closed geodesics of (N, h) with the periodic orbits of the geodesic flow on $T(N)$. The non-degenerate closed geodesics correspond then to the periodic orbits whose Poincaré maps have no eigenvalue equal to one. One then applies the proof of ([27], § 21), or of the stability theorem ([3], § 24) to obtain the desired conclusion.

(5.6) Let $\varphi_0 : (M, g_0) \rightarrow (N, h_0)$ be a harmonic map which factorises through the circle S^1 and whose only Jacobi fields are those induced by $\mathbf{I}(S^1)$. Combining (5.3) and (5.5), we see that the harmonic map will depend smoothly on a sufficiently small smooth deformation of the metrics.

The interest of this observation is that, if we exclude the trivial case of constant maps, that kind of degeneracy is the only possible one in the case $\text{Riem}^N < 0$. In particular, we can delete the condition (c) in remark (3.3).

6. Harmonic diffeomorphisms of surfaces

As an application, we give a new proof of the following result of Sampson [24] and Schoen-Yau [26] (their proofs involve a decomposition formula for the Laplacian of $|d\varphi|^2 : M \rightarrow \mathbf{R}$ and the maximum principle).

(6.1) *Theorem.* Let (M, g) and (N, h) be closed orientable surfaces of the same genus $p \geq 1$. Assume that the sectional curvature $\text{Riem}^N \leq 0$. If $\varphi : M \rightarrow N$ is a harmonic map of degree 1, then φ is a diffeomorphism.

Proof. Suppose first that $p \geq 2$, so that $\text{Riem}^N \neq 0$. Our hypothesis ensures that φ is homotopic to a diffeomorphism φ_0 . Let us consider a smooth path $(g_s)_{s \in I} \subset \mathbf{M}^\infty$ such that $g_0 = \varphi_0^* h$ and $g_1 = g$. The diffeomorphism $\varphi_0 : (M, g_0) \rightarrow (N, h)$ is then harmonic. Since $\text{Riem}^N \leq 0$ and $\neq 0$, for all $s \in I$ there is a unique harmonic map $\varphi_s : (M, g_s) \rightarrow (N, h)$ homotopic to φ_0 [10], [15]. By unicity, φ_0 is the given diffeomorphism and $\varphi_1 = \varphi$. Furthermore, $\nabla^2 E(\varphi_s)$ is non-degenerate, so that corollary (3.2) implies that $(\varphi_s)_{s \in I}$ is a continuous deformation of φ_0 to φ .

Let $S = \{t : J(\varphi_s) > 0 \text{ for } s \leq t\}$ where $J(\varphi_s)$ is the Jacobian of φ_s . Since a map of degree 1 with positive Jacobian is a diffeomorphism, it suffices to show that $S = I$. Therefore we remark that $0 \in S$ and show that S is open and closed.

That S is open follows from the compactness of M and the continuity of the above deformation. To show that S is closed, we consider a sequence $(t_i) \subset S$ such that $t_i \rightarrow t_\infty$. Then $J(\varphi_{t_\infty}) \geq 0$. By Wood's classification of singularities [32], φ_{t_∞} must be a branched covering. Since it is of degree 1, it is a diffeomorphism and $t_\infty \in S$.

The case $p = 1$ follows from standard methods (see [9], § 7).

By passing to the two-sheeted orientable coverings, we get the

(6.2) *Corollary.* Let (M, g) and (N, h) be non-orientable closed surfaces of the same genus $p \geq 1$. Assume $\text{Riem}^N \leq 0$. Any harmonic map $\varphi : (M, g) \rightarrow (N, h)$ homotopic to a diffeomorphism is a diffeomorphism.

Once in possession of (4.3), we can simplify somewhat the proof of the following result of Schoen-Yau [26] :

(6.3) *Theorem :* Let (M, g) be a compact surface with boundary and (N, h) a compact surface with $\text{Riem}^N \leq 0$. Suppose that $\psi : M \rightarrow N$ is a diffeomorphism onto $\psi(M)$. Assume that $\psi(\partial M)$ is convex with respect to $\psi(M)$. Then there is a (unique) harmonic map $\varphi : M \rightarrow N$ in the component of ψ in $C_{\psi|_{\partial M}}(M, N)$ such that $\varphi|_{M^0} \rightarrow \varphi(M^0)$ is a diffeomorphism.

As in the proof of theorem (6.1), we consider a deformation $(g_s)_{s \in I}$ of metrics on M such that $g_0 = \psi^* h$ and $g_1 = g$. For each g_s , Hamilton's existence theorem [14] ensures that ψ can be deformed with fixed boundary map to a harmonic map φ_s , unique by [15], [25], and such that $\varphi_s(M) \subset \psi(M)$. For suitable orientations this implies that $J(\varphi_s) \geq 0$ on ∂M . As in [26], we let $S = \{t : J(\varphi_s|_{M^0}) > 0 \text{ and } J(\varphi_s) \geq 0 \text{ for } s \leq t\}$. As above, S is closed; and we observe that the method of Schoen-Yau applies to this deformation to show that S is open : Indeed, use the complex structure of M to decompose $|d\varphi_s|^2 = |d'\varphi_s|^2 + |d''\varphi_s|^2$. Take $t_0 \in S$; then the continuous dependence theorem ensures that $|d'\varphi_s|^2 > 0$ on M for all t near t_0 . The proof of theorem 3.1 of [26] now applies, to show that $J(\varphi_t|_{M^0}) > 0$; we conclude that a neighbourhood of t_0 is contained in S .

7. Harmonic maps with non-negative Jacobian

We now examine briefly the possibility of extending the results of § 6 to other situations.

(7.1) It is unknown whether the curvature restriction in theorems (6.1) and (6.3) is necessary.

(7.2) Almost nothing is known about the singularities of harmonic maps between manifolds of dimensions > 2 . We note, however, that for $n \geq 3$, there are metrics g on the n -torus such that any harmonic map ϕ of (T^n, g) into a flat torus (T^n, g_0) has singularities. In particular, such a harmonic map homotopic to a diffeomorphism is not a diffeomorphism (although it might still be a homeomorphism).

Indeed, any such ϕ factors through the Albanese map α of (T^n, g) and its Jacobian $J(\alpha) = \omega_1 \wedge \dots \wedge \omega_n$, where $(\omega_i)_{1 \leq i \leq n}$ is a basis for the harmonic one-forms on (T^n, g) (for details see [9, § 7]). A study of the existence of metrics for which some harmonic one-form must have zeros has been made by Calabi [4].

(7.3) In view of theorem (6.1), we could ask whether a harmonic map between surfaces which is homotopic to a surjective branched covering is itself a branched covering. That this is not the case, even if the target has negative curvature, is shown by the

(7.4) *Example*: Let N be a compact orientable surface without boundary. For any Riemannian metric h on N , there exist a Riemann surface M and a harmonic map $\phi : M \rightarrow (N, h)$ which is homotopic to a surjective branched covering but is not a branched covering. In particular, if N has negative curvature then the harmonic map is unique in its homotopy class, which therefore contains no harmonic branched covering.

On the other hand, note that if N is a sphere, certain homotopy classes can contain two harmonic maps, only one of which is a branched covering [18].

We first construct the example in the case where genus $N = q \geq 1$.

Take any Riemann surface P of genus $q + 1$. We represent P by the polygon $a_1 b_1 a_1^{-1} b_1^{-1} \dots a_{q+1} b_{q+1} a_{q+1}^{-1} b_{q+1}^{-1}$ with the usual identifications. Its fundamental group $\pi_1(P)$ is then generated by $(a_i, b_i)_{1 \leq i \leq q+1}$ with the single relation $\prod_{i=1}^{q+1} a_i b_i a_i^{-1} b_i^{-1} = 1$. We represent N similarly, using $(c_i, d_i)_{1 \leq i \leq q}$. The homomorphism $\pi_1(P) \rightarrow \pi_1(N)$ defined by $a_i \rightarrow c_i$, $b_i \rightarrow d_i$ ($1 \leq i \leq q$), $a_{q+1} \rightarrow 1$ and $b_{q+1} \rightarrow 1$ determines a homotopy class of maps $P \rightarrow N$ of degree 1. By [17] that class contains a harmonic map ψ .

Let M be the surface of genus $M = 2q + 1$ obtained by joining cross-wise two copies of P along a closed curve cutting the handle of P represented by (a_{q+1}, b_{q+1}) . M admits a complex structure such that the covering map $\beta : M \rightarrow P$ is holomorphic.

The composition $\psi \circ \beta : M \rightarrow N$ is then harmonic and is not a branched covering. Since its degree is two and it induces an epimorphism on fundamental groups, it is homotopic to a branched covering by a result of Edmonds [7]. That is a consequence of a theorem stating (in his terminology) that any map $M \rightarrow N$ is homotopic to a pinch followed by a branched covering. We remark that our map ϕ is the composition of a covering and a pinch, but taken in opposite order.

If genus $N = 0$, the existence of certain harmonic maps from a surface of genus ≥ 3 to N which are homotopic to a branched covering but are not branched coverings is established in [18].

8. Examples

The results of §§ 3–5 affirm that a sufficiently small smooth deformation of metrics induces a smooth variation of the associated harmonic maps, provided that the curvature of N or the initial harmonic map satisfies some restrictive hypothesis. By means of examples, we shall illustrate now the nature of these restrictions.

(8.1) Sampson has shown us the necessity of making some sort of hypothesis by studying the behaviour of a closed geodesic on a flat cylinder, deformed into a conical shape. His example can be made more explicit as follows.

Let N be a piece of cylindrical surface $(a, b) \times S^1$ equipped in local coordinates $(u^1, u^2) \in (a, b) \times \mathbf{R}$ with the metric $h_{11} = 1, h_{22} = f(u^1), h_{12} = 0$. At a point (u^1, u^2) , the curvature of N is given by

$$K = \frac{1}{4} \frac{f'^2}{f^2} - \frac{1}{2} \frac{f''}{f}.$$

For any point $c \in (a, b)$, the tension of the map $\varphi : S^1 \rightarrow N$ defined by $\varphi^1(x) = c$, $\varphi^2(x) = x$ is given by $\tau^1(\varphi) = -\frac{1}{2}f'$, $\tau^2(\varphi) = 0$. Therefore, such a map is a geodesic iff $f'(c) = 0$. If $f'(c) < 0$, the set $\{(u^1, u^2) \mid u^1 > c\}$ is the concave side of the curve φ .

Suppose now that f is constant, so that N is a piece of flat cylinder. For any $c \in (a, b)$, the map $\varphi : S^1 \rightarrow N$ defined above is a geodesic and the vector field along φ defined by $v^1(x) = \text{const.}$, $v^2(x) = 0$ is a Jacobi field which is not a Killing field of S^1 . Therefore, φ does not satisfy the hypothesis of proposition (5.5).

If we now modify the metric on N in such a way that f becomes decreasing, we observe that N does not carry any non-trivial closed geodesic, so that proposition (5.5) fails. Indeed, the image of any map $\varphi : S^1 \rightarrow N$ is situated on the concave side of the line $u^1 = \min \varphi^1$. By Sampson's maximum principle [24], φ cannot be harmonic.

In this example the degeneracy of φ is due to the presence of a continuous family of closed geodesics, obtained by varying c in (a, b) . That proposition (5.5) can fail also for an isolated geodesic (up to rotation on S^1) can be seen by defining the initial metric by a monotone function f such that $f'(c) = 0 = f''(c)$; for example $f = 1 + (u^1 - c)^3$. Then the curve $u^1 = c$ is a geodesic, and the vector field along the curve defined by $v^1 = \text{const.}$, $v^2 = 0$ is a Jacobi field.

For the perturbed metric determined by $f = 1 + (u^1 - c)^3 + \epsilon(u^1 - c)$, $\epsilon > 0$, there is no geodesic in a neighbourhood of $u^1 = c$, by the maximum principle.

In these examples, we have considered only a piece of surface N . Using partitions of unity, we can extend N to form a complete Riemannian surface.

When that surface is compact, there is always a closed geodesic homotopic to a given curve; thus the examples show that a smooth deformation of metrics does not determine a continuous variation of harmonic maps, although the existence of such maps is not in question here.

For N non-compact, however, an arbitrarily small deformation of metric could make a harmonic map disappear altogether. Indeed, we can deform the flat metric on an infinite cylinder to a metric such that $f' < 0$. Then the modified surface does not carry any closed geodesic.

(8.2) In all the above examples, the harmonic map under consideration is a geodesic, so that the rotation group of S^1 always induces a degeneracy of the Hessian of E . Therefore, these examples pertain to the situation considered in § 5.

It is a simple matter to adapt the examples to suppress that Killing degeneracy. Indeed, it suffices in all the constructions to replace the maps $S^1 \rightarrow (a, b) \times S^1$ by maps $M \rightarrow (a, b) \times M$ defined analogously, where M is a compact manifold with discrete isometry group. In that way, one sees that the hypothesis of non-degeneracy of the Hessian in theorem (3.1) is essential.

(8.3) In relation to deformations of the metric of M , we have the following partial result. Let M be the flat torus $\mathbb{R}/2\pi d\mathbb{Z} \times \mathbb{R}/2\pi\mathbb{Z}$ equipped with Euclidean coordinates (x, y) and N the piece of surface defined as above with $u^1 \in (-b, b)$ and $f = (b + A)^2 - (u^1)^2$ (where $A > 0$). It was observed in [18] that a map $\varphi: M \rightarrow N$ of the form $\varphi(x, y) = (\psi(x), y)$ is harmonic iff ψ satisfies the pendulum equation

$$\frac{d^2 \psi}{dx^2} + \psi = 0.$$

For $d = 1$, the "length" of the torus coincides with the period of the pendulum and there is a continuous family of harmonic maps. For $d \neq 1$, the only harmonic map of the prescribed form is given by $\psi = 0$, so that we do not have a smooth variation of the harmonic maps in this restricted class.

(8.4) It might be tempting to try to prove existence theorems for harmonic maps, by deforming a metric for which existence is known to another one. The following examples show that this cannot always work:

(8.5) Let M be a closed orientable surface of genus 3 and (N, h) a 2-sphere. It is shown in [11] that there is a harmonic map $\varphi: (M, g) \rightarrow (N, h)$ of degree 3 iff g is a Hermitian metric associated to a non-hyperelliptic complex structure on M . Call $\mathbf{E}$ the set of such metrics. Now $\mathbf{E}$ is open in $\mathbf{M}$; although $\nabla^2 E(\varphi)$ is degenerate, we see that for any $g \in \mathbf{E}$, a small deformation of g will preserve the existence of a harmonic map. Along a path leading from $\mathbf{E}$ to $\mathbf{M} \setminus \mathbf{E}$, however, existence will not be preserved. With respect to this observation, we stress that theorem (3.1) concerns only small variations of metrics, and leads to the global statement of corollary (3.2) only if existence and unicity of the harmonic maps are assumed.

A similar disappearance occurs with respect to variations of Dirichlet data; see ([9], 12.6).

(8.6) Let N be the infinite cylinder on which we define a metric h_ϵ as above, using the function $f = e^{-u^1} + \epsilon(u^1)^2$. The curvature of (N, h_ϵ) is negative for all ϵ and a negative constant when $\epsilon = 0$. For any $\epsilon > 0$, (N, h_ϵ) carries the closed geodesic $u^1 = c$, where c is the solution of $e^{-c} = 2\epsilon c$. For $\epsilon = 0$, the maximum principle implies that (N, h_0) carries no closed geodesic. Note that (N, h_0) does not satisfy the growth condition of remark (3.3).

(8.7) Finally, we recall [28] that if $n \geq 3$, $M = S^n$ is the Euclidean sphere and N is an ellipsoid of revolution of dimension n , one of whose axis is A times the others, then there is a harmonic map $\varphi : M \rightarrow N$ which is a suspension of the identity on S^{n-1} if A is small, but not if A is large. What happens along a deformation of the metric of N leading from one value of A to the other is not known.

References

- [1] Abraham R 1962 *Lectures of Smale on differential topology*. Columbia University
- [2] Abraham R 1970 Bumpy metrics; *Proc. Symp. Pure. Math.* **14** 1-3
- [3] Abraham R and Robbin J 1967 *Transversal mappings and flows* (New York : Benjamin)
- [4] Calabi E 1969 An intrinsic characterisation of harmonic one-forms; *Global Analysis* (Papers in honour of K Kodaira) (Princeton : University Press and 1969 University of Tokyo Press) 101-117
- [5] Dieudonne J A 1960 *Foundations of modern analysis* (New York : Academic Press)
- [6] Earle C J and Eells J 1969 A fibre bundle description of Teichmüller theory; *J. Diff. Geom.* **3** 19-43
- [7] Edmonds A L 1980 Deformations of maps to branched coverings in dimension two (to appear)
- [8] Eells J 1966 A setting for global analysis; *Bull. Am. Math. Soc.* **72** 751-807
- [9] Eells J and Lemaire L 1978 A report on harmonic maps; *Bull. London Math. Soc.* **10** 1-68
- [10] Eells J and Sampson J H 1964 Harmonic mappings of Riemannian manifolds; *Am. J. Math.* **86** 109-160
- [11] Eells J and Wood J C 1976 Restrictions on harmonic maps of surfaces; *Topology* **15** 263-266
- [12] Gilbarg D and Trudinger N S 1977 *Elliptic partial differential equations of second order* (Grundlehren 224 : Springer Verlag)
- [13] Goldstein R A 1972 Stability of the boundary value problem for harmonic mappings; *J. Math. Anal. Appl.* **39** 346-359
- [14] Hamilton R S 1975 *Harmonic maps of manifolds with boundary*. Springer lecture notes 471
- [15] Hartman P 1967 On homotopic harmonic maps; *Can. J. Math.* **19** 673-687
- [16] Hörmander L 1964 *Linear partial differential operators* (Grundlehren : Springer Verlag)
- [17] Lemaire L 1975 Applications harmoniques de surfaces; *CR Acad. Sci. Paris* **A280** 897-899
- [18] Lemaire L 1978 Harmonic nonholomorphic maps from a surface to a sphere; *Proc. Am. Math. Soc.* **71** 299-304
- [19] Lichnerowicz A 1961 Propagateurs et commutateurs en relativite generale; *IHES Publ.* **10**
- [20] Mazet E 1973 La formule de la variation seconde de l'energie au voisinage d'une application harmonique; *J. Diff. Geom.* **8** 279-296
- [21] Morrey C B 1966 *Multiple integrals in the calculus of variations*. (Grundlehren : Springer Verlag)
- [22] Morse M 1934 The calculus of variations in the large; *Am. Math. Soc. Colloq. Publ.* **18**
- [23] Royden H L 1960 The analytic approximation of differentiable mappings; *Math. Ann.* **139** 171-179
- [24] Sampson J H 1978 Some properties and applications of harmonic mappings; *Ann. Ec. Norm. Sup.* **11** 221-228
- [25] Schoen R M 1977 *Existence and regularity of some geometric variational problems*; Thesis, Stanford Univ.
- [26] Schoen R M and Yau S T 1978 On univalent harmonic maps between surfaces; *Invent. Math.* **44** 265-278
- [27] Siegel C L and Moser J K 1971 *Lectures on classical mechanics* (Grundlehren : Springer Verlag)

- [28] Smith R T 1975 Harmonic mappings of spheres; *Am. J. Math.* **97** 364–385
- [29] Smith R T 1975 The second variation formula for harmonic mappings; *Proc. Am. Math. Soc.* **47** 229–236
- [30] Visik M I 1951 On strongly elliptic systems of differential equations; *Math. Sbornik* **29** 615–676
- [31] Whitney H 1934 Analytic extension of differentiable functions defined on closed sets; *Trans. Am. Math. Soc.* **36** 63–89
- [32] Wood J C 1977 Singularities of harmonic maps and applications of the Gauss-Bonne formula; *Am. J. Math.* **99** 1329–1344

J. Eells

Department of Mathematics
University of Warwick
Coventry CV4 7AL (UK)

L. Lemaire

Institut des Hautes Etudes Scientifiques
35 Rue de Chartres
91440 Bures-sur-Yvette, France

Also

Université Libre de Bruxelles
(Belgium)



Curvature and the heat equation for the de Rham complex

By

PETER B GILKEY*

1. Introduction

Let M be a compact Riemannian manifold without boundary, of dimension m . The Chern-Gauss-Bonnet theorem [2] gives a formula for the Euler-Poincaré characteristic $\chi(M)$ in terms of curvature. In 1967, McKean and Singer [8] conjectured that this theorem could be proved by heat equation methods. In 1971, Patodi [10] was able to prove this conjecture using a complicated cancellation argument. He was able to generalise his methods to give a proof of the Hirzebruch signature formula and of the Riemann-Roch formula for Kaehler manifolds as well using heat equation methods.

In 1972, the present author gave a different proof of these results using the techniques of invariance theory as a substitute for Patodi's cancellation lemma [4]. In this paper, we use these techniques to compute some of the higher order terms in the asymptotic expansion of the heat equation for the de Rham complex.

1. Statement of results

Let G denote the metric tensor of M . Choose a local orthonormal frame $e = (e_1, \dots, e_m)$ for the tangent space $T(M)$. We use the metric to identify $T(M)$ and $T^*(M)$ so we may also regard e as a local frame for $T^*(M)$. Let $R_{ijkl}(x, G, e)$ be the curvature tensor of the Levi-Civita connection on $T(M)$ at the point $x \in M$. We have chosen the sign convention so $R_{1212} = -1$ on the unit sphere in R^3 .

We adopt the Einstein convention of summing over repeated indices unless otherwise indicated. Let $I = (i_1, \dots, i_m)$ and $J = (j_1, \dots, j_m)$ be two ordered collections of indices. Define

$$\varepsilon(I, J) = G(e_{i_1} \wedge \dots \wedge e_{i_m}, e_{j_1} \wedge \dots \wedge e_{j_m}).$$

If the two collections of indices I and J consist of distinct indices, ε is the sign of the permutation which transforms I to J . Otherwise ε is zero.

* Research partially supported by NSF grant MCS76-23465 and by Sloan Foundation Grant BR-1687.

The integrand E_m of the Chern-Gauss-Bonnet theorem vanishes if m is odd. If m is even, define

$$E_m(x, G) = \frac{1}{(-8\pi)^{\frac{1}{2}m} (\frac{1}{2}m)!} \sum_{I, J} \varepsilon(I, J) R_{i_1 i_2 i_1 i_2} \dots R_{i_{m-1} i_m i_{m-1} i_m}(x, G, e).$$

$E_m(x, G)$ is a scalar invariant of G which does not depend on the choice of e . Let $|d\text{vol}|(x)$ be the Riemannian measure on M . (We use the notation $|d\text{vol}|$ to emphasise that this is a measure rather than an m -form and is therefore independent of the orientation.) The Chern-Gauss-Bonnet theorem states:

$$\chi(M) = \int_M E_m(x, G) |d\text{vol}|(x).$$

We now discuss the heat equation proof of this theorem found by Patodi.

$$\text{Let } D_p^m = (d\delta + \delta d): C^\infty \Lambda^p T^*M \rightarrow C^\infty \Lambda^p T^*M,$$

be the Laplacian acting on the space of smooth p -forms. (We suppress the dependence of D_p^m on the metric G for notational convenience.) The fundamental solution of the heat equation is $\exp(-tD_p^m)$. This extends to an infinitely smoothing operator mapping $L^2(\Lambda^p T^*M) \rightarrow C^\infty \Lambda^p T^*M$ for $t > 0$. This operator has a smooth kernel function we will denote by $K(t, x, y, D_p^m)$; K is bundle morphism-valued. K vanishes to infinite order for $x \neq y$ as $t \rightarrow 0^+$. If $x = y$, there is an asymptotic expansion of the form:

$$K(t, x, x, D_p^m) \sim t^{-\frac{1}{2}m} \sum_{n=0}^{\infty} t^n E_n(x, D_p^m) \text{ as } t \rightarrow 0^+.$$

$E_n(x, D_p^m)$ is a smooth endomorphism of the bundle of p -forms. It has been computed for $n = 0, 1, 2, 3$ in [5, 7].

The surprising fact is that the coefficients E_n are smooth local invariants of the Riemannian metric which can be expressed in terms of the covariant derivatives of the curvature tensor. We define:

$$a_n(x, D_p^m) = \text{Tr } E_n(x, D_p^m) \text{ and } a_n(D_p^m) = \int_M a_n(x, D_p^m) |d\text{vol}|(x).$$

(We have omitted the normalising constant of $(4\pi)^{-\frac{1}{2}m}$ in order to simplify later expressions and therefore our definition differs slightly from that found elsewhere in the literature.) It follows that:

$$\begin{aligned} \text{Tr}_{L^2} \exp(-tD_p^m) &= \int_M \text{Tr}_{\Lambda^p} K(t, x, x, D_p^m) |d\text{vol}|(x) \\ &\sim t^{-\frac{1}{2}m} \sum_{n=0}^{\infty} t^n \int_M a_n(x, D_p^m) |d\text{vol}|(x) \\ &\sim t^{-\frac{1}{2}m} \sum_{n=0}^{\infty} t^n a_n(D_p^m). \end{aligned}$$

This shows that the invariants $a_n(D_p^m)$ are determined by the spectrum of the Laplacian. $a_n(D_p^m)$ was calculated by Patodi for all (m, p) and for $n = 0, 1, 2$ [9]. His results have been extended by Sakai and the author to the case $n = 3$ [11, 5].

We can use the local invariants $a_n(x, D_p^m)$ to obtain a local formula for $\chi(M)$. There is a purely formal algebraic lemma, which is valid for any elliptic complex, that shows [3, 4]

$$\begin{aligned} \text{index}(d + \delta) &= \chi(M) = \sum_p (-1)^p \dim N(D_p^m) \\ &= \sum_p (-1)^p \text{Tr}_{L^2} \exp(-tD_p^m) \end{aligned}$$

for any $t > 0$.

Let $a_n^m(x, G) = \sum_p (-1)^p a_n(x, D_p^m)$.

The asymptotic series for $\exp(-tD_p^m)$ together with this algebraic fact proves:

Lemma 1.1

$$\int_M a_n^m(x, G) |d\text{vol}|(x) = \begin{cases} 0 & \text{if } 2n \neq m \\ \chi(M) & \text{if } 2n = m \end{cases}.$$

This proves $\chi(M) = 0$ if m is odd. Using a local version of Poincaré duality one can prove the local invariants $a_n^m(x, G) \equiv 0$ as well if m is odd. We shall therefore restrict to the case of even-dimensional manifolds in much of what follows.

Lemma 1.1 led McKean and Singer [8] to conjecture that some fantastic cancellation might take place and that $a_n^m(x, G)$ would be the integrand of the Chern-Gauss-Bonnet theorem [2] if $2n = m$. They were able to verify their conjecture for the special case $m = 2$, but the higher dimensional case was left open. Patodi [10] proved their conjecture by showing:

Theorem 1.2

(a) $a_n^m(x, G) = 0$ if $2n < m$ or if m is odd

(b) $a_n^m(x, G) = E_m(x, G)$ if $2n = m$.

It is known that $a_n^m(x, G) \neq 0$ if $2n > m$ and if m is even [7]. It is also known that there exists a functorial 1-form valued invariant $Q_n^m(x, G)$ so $a_n^m(x, G) = \text{divergence } Q_n^m(x, G)$ for $2n > m$ [6]. However, aside from the case $m = 2$, little else is known about the higher invariants $a_n^m(x, G)$ for $2n > m$. In this paper, we compute the next invariant beyond the Euler form. Let ‘;’ denote multiple covariant differentiation and define:

$$\Phi_m(x, G) = \frac{1}{(-8\pi)^{\frac{1}{2}m} (\frac{1}{2}m - 1)!} \sum_{I, J, k} \varepsilon(I, J) R_{i_1 i_2 i_3 k; k i_2} R_{i_3 i_4 i_5 i_6} \dots R_{i_{m-1} i_m i_{m-1} i_m}.$$

Theorem 1.3

If $2n = m + 2$

then $a_n^m(x, G) = \frac{1}{12} (E_m(x, G))_{;kk} + \frac{1}{6} \Phi_m(x, G)$.

In § 2, we shall discuss the space of invariants $\mathcal{K}_{2n}^m$ to which the local invariants $a_n^m(x, G)$ belong. In theorem 2.5, we give a basis for the vector spaces $\mathcal{K}_{2n}^m$ for $2n = m$, $2n = m + 1$, and for $2n = m + 2$. This will lead to the proof of theorem 1.2 using invariance theory that we gave earlier [4]. In § 3 we combine the results of theorem 2.5 with some formulas derived in [5] to complete the proof of theorem 1.3. We also discuss some open questions which relate to the invariants a_n^m for $2n > m + 2$.

2. Invariance theory

Before proving theorem 1.3, we must discuss the space of invariants to which the $a_n^m(x, G)$ belong. Let $X = (x_1, \dots, x_m)$ be a system of local coordinates on M and let

$$g_{ij}(X, G)(x) = G(\partial/\partial x_i, \partial/\partial x_j).$$

Let $\alpha = (\alpha(1), \dots, \alpha(m))$ be a multi-index.

Let $|\alpha| = \alpha(1) + \dots + \alpha(m)$.

Let $d_X^\alpha = (\partial/\partial x_1)^{\alpha(1)} \dots (\partial/\partial x_m)^{\alpha(m)}$ be ordinary partial differentiation. We define

$$g_{ij|\alpha}(X, G)(x) = d_X^\alpha g_{ij}(X, G)(x).$$

We shall regard the $\{g_{ij|\alpha}\}$ as formal variables which represent the jets of the metric. We can evaluate the $g_{ij|\alpha}$ variables on the metric once the coordinate system has been chosen, but these variables have no invariant or tensorial meaning.

Let $\mathcal{P}^m$ denote the polynomial algebra in the $\{g_{ij|\alpha}\}$ variables for $|\alpha| > 0$ with coefficients which are smooth functions of the $\{g_{ij}\}$ variables. If $P \in \mathcal{P}^m$, let $P(X, G)(x)$ be the evaluation of P on the metric G relative to the coordinate system X . We say that P is invariant if $P(X, G)(x) = P(Y, G)(x)$ for any two coordinate systems X and Y defined near x . If P is invariant, we will denote this common value by $P(G)(x)$. The scalar curvature $K = -\frac{1}{2} R_{ijij}$ and the integrand of the Chern-Gauss-Bonnet theorem E_m are examples of invariant elements of $\mathcal{P}^m$.

We define a grading on $\mathcal{P}^m$ as follows : let $\text{ord}(g_{ij|\alpha}) = |\alpha|$.

If $A = a(g_{ij}) g_{i_1 j_1 |\alpha_1} \dots g_{i_r j_r |\alpha_r}$,

is a monomial of P , let $\text{ord}(A) = |\alpha_1| + \dots + |\alpha_r|$. If all the monomials of $P \in \mathcal{P}^m$ are of order n , we shall say P is homogeneous of order n . The following lemma was proved in [4] and is a simple consequence of dimension analysis :

Lemma 2.1

Let $P \in \mathcal{P}^m$ be invariant and homogeneous of order n . Let $c > 0$, then

$$P(c^2 G)(x) = c^{-n} P(G)(x).$$

This relates the order of a polynomial P to the effect on its value of a constant dilation of the metric. Let $\mathcal{P}_n^m$ denote the vector space of all invariant polynomials of $\mathcal{P}^m$ which are homogeneous of order n . We will show later $\mathcal{P}_n^m = 0$ if n is odd. If $P \in \mathcal{P}^m$, we may decompose $P = P_0 + \dots + P_k$ into homogeneous polynomials P_n of order n . If P is invariant, it is not difficult to show each of the

P_n are invariant separately. Therefore the space of all invariant polynomials of $\mathcal{P}^m$ may be decomposed into the direct sum of the spaces $\mathcal{P}_n^m$.

We fix a basepoint $x_0 \in M$ and normalise the coordinate system. We will assume $g_{ij}(X, G)(x_0) = \delta_{ij}$ is the Kronecker index and that all the first derivatives of the metric vanish at x_0 . By restricting to such coordinate systems, we can regard $\mathcal{P}^m$ as a polynomial algebra in the $\{g_{ij/a}\}$ variables for $|a| \geq 2$ with constant coefficients. The coordinate frame $(\partial/\partial x_i)$ is an orthonormal frame for $T(M)$ at x_0 . Although this normalisation is useful for proving theorems, it is important for actual applications to know that the invariants of $\mathcal{P}_n^m$ can be evaluated in any coordinate system.

Let $R_{ijkl, \dots}$ denote the components of the covariant derivatives of the curvature tensor. R_{ijkl} is linear in the second derivatives of the metric and quadratic in the first derivatives of the metric with coefficients which are smooth functions of the metric tensor variables g_{ij} . $R_{ijkl} \in \mathcal{P}^m$ is homogeneous of order 2; each additional covariant derivative increases the order by 1. This shows $E_m \in \mathcal{P}_m^m$ for all m .

By using the functorial properties of the heat equation, it can be shown that $a_n^m(x, G)$ is homogeneous of order $2n$. Since a_n^m is defined using the kernel of the heat equation, $a_n^m(X, G)$ does not depend on the choice of the coordinate system so $a_n^m \in \mathcal{P}_{2n}^m$. This property alone is not sufficient to determine a_n^m and to complete the proofs of theorems 1.2 and 1.3. We need an additional property which we describe as follows. Let $M_1^{m_1}$ and $M_2^{m_2}$ be two Riemannian manifolds with metrics G_1 and G_2 . Let $M = M_1 \times M_2$ with the product metric $G = G_1 \times G_2$. In [4], we used the formal properties of the de Rham complex under isometric product to show that :

Lemma 2.2

If $m = m_1 + m_2$,

then $a_n^m(G_1 \times G_2)(x_1, x_2) = \sum_{p+q=n} a_p^{m_1}(G_1)(x_1) a_q^{m_2}(G_2)(x_2)$.

Any two metrics on S^1 are locally isometric. Let $M_2 = S^1$. Since this is a homogeneous space, a_p^1 is constant. Since $\chi(S^1) = 0$, lemma 1.1 implies $a_p^1 \equiv 0$. Thus if $M = M_1 \times S^1$ with the product metric, $a_n^{m_1}(G_1 \times 1)(x_1, \theta) = 0$ for every $\theta \in S^1$. We shall exploit this property to prove theorems 1.2 and 1.3.

More generally, let $P \in \mathcal{P}_n^m$ be invariant and let $M_1^{m_1}$ have metric G_1 . Let $M = M_1^{m_1} \times S^1$ have product metric $G_1 \times 1$. The circle acts on M by isometries so $P(G_1 \times 1)(x, \theta)$ is independent of θ ; we denote this common value by $rP(G_1)(x_1)$. r defines an algebra morphism from the invariant polynomials of $\mathcal{P}^m$ to the invariant polynomials of $\mathcal{P}^{m-1}$. By using the characterisation of order given in lemma 2.1, we observe that r preserves the order and induces a map $r : \mathcal{P}_n^m \rightarrow \mathcal{P}_n^{m-1}$. Let $\mathcal{K}_n^m$ be the kernel of r ; then $r(a_n^m) = 0$ so $a_n^m \in \mathcal{K}_{2n}^m$.

r can also be described by using Weyl's theorem [12] on the invariants of the orthogonal group. Although our discussion does not depend on this result, we include this description for the sake of completeness. Further details may be found in Atiyah *et al* [1]. In normal coordinates, we can express the jets of the metric in terms of the $R_{ijkl, \dots}$ variables so $P \in \mathcal{P}_n^m$ can be viewed as an invariant

expression in the $R_{ijkl;...}$ variables. By Weyl's theorem, this implies P can be expressed in terms of contractions of indices. For example, the scalar curvature $K = -\frac{1}{2} R_{ijij}$. The range of the indices summed over ranges from 1 through m if $P \in \mathcal{P}_n^m$; $rP \in \mathcal{P}_n^{m-1}$ is defined by summing over indices which range from 1 through $m-1$; we set to zero any of the $R_{ijkl;...}$ variables which involve the last index m . Such expressions in terms of contractions of indices are universal in the sense that the dependence on the dimension m is not explicit. The spaces $\mathcal{K}_n^m$ measure the additional relations which are introduced when the dimension is reduced by 1.

From this description of the map r , it is clear that r is always surjective. We proved earlier [4] that :

Theorem 2.3

(a) $\mathcal{K}_n^m = 0$ if $n < m$

(b) $\mathcal{K}_m^m$ is one-dimensional and is spanned by E_m .

We can derive theorem 1.2 from theorem 2.3 as follows : we observed above that $a_n^m \in \mathcal{K}_{2n}^m$. Therefore $a_n^m = 0$ if $2n < m$. If $2t = m$, $a_t^m = c(m) E_m$ for some universal constant $c(m)$. The constant $c(m)$ can be evaluated by integration over the test manifold $M_0 = S^2 \times \dots \times S^2$ with the standard product metric G_0 . The normalising constants in the definition of E_m were chosen so

$$\int_{M_0} E_m(x, G_0) |d\text{vol}|(x) = 2^t.$$

$$\text{Since } \int_{M_0} a_t^m(x, G) |d\text{vol}|(x) = \chi(M) = 2^t,$$

this implies $c(m) = 1$ which proves theorem 1.2.

We will prove theorem 1.3 by generalising theorem 2.3 appropriately. Let $2t = m$ and define :

$$\psi_{m+1} = \frac{1}{(-8\pi)^t (t-1)!} \sum_{I,J} \varepsilon(I, J) \sum_k R_{k i_1 j_1 i_2} R_{k i_2 j_2 i_3} R_{k i_3 j_3 i_4} \dots$$

$$R_{i_m i_{m+1} j_m j_{m+1}} \in \mathcal{K}_{m+2}^{m+1},$$

$$\Theta_m^1 = \frac{1}{(-8\pi)^t (t-1)!} \sum_{I,J} \varepsilon(I, J) \sum_k R_{k i_1 j_1 i_2 j_3} R_{k i_3 j_3 i_4 j_5} \dots$$

$$R_{i_{m-1} i_m j_{m-1} j_m} \in \mathcal{K}_{m+2}^m,$$

$$\Theta_m^2 = \frac{1}{(-8\pi)^t (t-2)!} \sum_{I,J} \varepsilon(I, J) \sum_k R_{k i_1 j_1 i_2 j_3} R_{k i_3 j_3 i_4 j_5} R_{k i_5 j_5 i_6 j_7} \dots$$

$$R_{i_{m-1} i_m j_{m-1} j_m} \in \mathcal{K}_{m+2}^m,$$

$$\Theta_m^3 = \frac{1}{(-8\pi)^t (t-1)!} \sum_{I,J} \varepsilon(I, J) \sum_{k,l} R_{k i_1 j_1 i_2} R_{l i_3 j_3 i_4} R_{l i_5 j_5 i_6} \dots$$

$$R_{i_{m-1} i_m j_{m-1} j_m} \in \mathcal{K}_{m+2}^m,$$

$$\Theta_m^4 = \frac{1}{(-8\pi)^t t!} \Sigma_{I,J} \varepsilon(I, J) \Sigma_{k,l} R_{k|kl} R_{i_1 i_2 j_1 j_2} R_{i_3 i_4 j_3 j_4} \dots$$

$$R_{i_{m-2} i_m j_{m-1} j_m} \in \mathfrak{K}_{m+2}^m,$$

$$\Theta_m^5 = \frac{1}{(-8\pi)^t (t-1)!} \Sigma_{I,J} \varepsilon(I, J) \Sigma_{k,l} R_{k|li_1 i_2} R_{k|j_1 j_2} \dots R_{i_{m-1} i_m j_{m-1} j_m} \in \mathfrak{K}_{m+2}^m,$$

$$\Theta_m^6 = \frac{1}{(-8\pi)^t (t-3)!} \Sigma_{I,J} \varepsilon(I, J) \Sigma_{k,l} R_{k i_1 j_1 j_2} R_{k j_2 i_2 j_3} R_{l i_3 j_3 j_4} R_{l j_4 i_4 j_5} \\ \times R_{i_1 i_2 j_1 j_2} \dots \in \mathfrak{K}_{m+2}^m.$$

We define $\Psi_{m+1} = \Theta_m^4 = 0$ if m is odd. Similarly $\theta_2^6 = \theta_4^6 = \theta_2^4 = 0$.

Lemma 2.4

- (a) $\Psi_{m+1} \neq 0$ if $m+1 \geq 3$,
- (b) $\{\Theta_m^i\}_{i=1}^6$ are linearly independent in $\mathfrak{K}_{m+2}^m$ for $m \geq 6$,
- (c) $\{\Theta_4^i\}_{i=1}^5$ are linearly independent in $\mathfrak{K}_6^4$
- (d) $\{\Theta_2^1, \Theta_2^3\}$ are linearly independent in $\mathfrak{K}_4^2$,

Proof: Let $P \in \mathcal{P}^m$ and let A be a monomial. We define $c(A, P)$ to be the coefficient of A in P . We will say that A is a monomial of P if $c(A, P) \neq 0$. The functions $c(A, P)$ are linear functionals on the spaces $\mathcal{P}_a^n$.

Let $g_{ij/k_1 \dots k_s} = (\partial/\partial x_{k_1}) \dots (\partial/\partial x_{k_s}) g_{ij}$.

This is another notation for the jets of the metric that will be convenient to have in what follows.

Since $g_{ij}(x_0) = \delta_{ij}$ and $g_{ij/k}(x_0) = 0$,

the curvature tensor is given by :

$$R_{ijkl} = \frac{1}{2} (g_{ik/jl} + g_{jl/ik} - g_{il/jk} - g_{jk/il}) + 0(x - x_0).$$

Let $B_{m+1} = g_{11/22} g_{33/22} g_{44/55} \dots g_{mm/m+1, m+1}$ for $m+1 \geq 3$,

$$C_m^1 = g_{11/2222} g_{33/44} g_{55/66} \dots g_{m-1, m-1/m m} \text{ for } m \geq 2,$$

$$C_m^2 = g_{11/222} g_{33/244} g_{55/56} \dots g_{m-1, m-1/m m} \text{ for } m \geq 4,$$

$$C_m^3 = g_{11/22} g_{33/22} g_{44/22} g_{55/66} \dots g_{m-1, m-1/m m} \text{ for } m \geq 2,$$

$$C_m^4 = g_{11/22} g_{33/22} g_{44/11} g_{55/66} \dots g_{m-1, m-1/m m} \text{ for } m \geq 4,$$

$$C_m^5 = g_{11/22} g_{11/22} g_{33/44} g_{55/66} \dots g_{m-1, m-1/m m} \text{ for } m \geq 4,$$

$$C_m^6 = g_{33/22} g_{44/22} g_{55/11} g_{66/11} g_{77/88} \dots g_{m-1, m-1/m m} \text{ for } m \geq 6.$$

It is clear from the definition that if the metric is flat in any index that $\Psi_{m+1} = 0$ and $\Theta_m^i = 0$. Therefore

$$\Psi_{m+1} \in \mathfrak{K}_{m+2}^{m+1} \text{ and } \Theta_m^i \in \mathfrak{K}_{m+2}^m \text{ for } i = 1, \dots, 6,$$

We will complete the proof of lemma 2.4 by showing that $c(B_{m+1}, \Psi_{m+1}) \neq 0$ and that the matrix $c(C_m^i, \theta_m^j)$ is upper triangular with non-zero diagonal entries. The calculation is somewhat technical and may be skipped without any loss of continuity.

We compute $c(B_{m+1}, \Psi_{m+1})$ to prove (a) : the index '2' appears four times in the monomial B_{m+1} . Since any index appears once in each of the collections I and J , only those terms with $k=2$ in the sum defining Ψ_{m+1} can involve the monomial B_{m+1} . By multiplying by 4^t and using the symmetries of the curvature tensor, we may further restrict the sum to range over (I, J) with $i_{2k} < i_{2k+1}$ and $j_{2k} < j_{2k+1}$ for $k \geq 1$. Since $B_{m+1} = g_{11/22} g_{33/22} \dots$, either $i_1 = 1$ or $i_1 = 3$. Since these two cases are equivalent, we multiply by 2 and suppose $i_1 = 1$. Only those terms with $(i_1, i_2, i_3, j_1, j_2, j_3) = (1, 2, 3, 3, 1, 2)$ can involve B_{m+1} . Since the remainder of B_{m+1} only involves $g_{2a, 2a/2a+1, 2a+1}$, $i_{2k} = j_{2k}$ and $i_{2k+1} = j_{2k+1} = i_{2k+1}$ for $k \geq 2$. This implies the indices $(i_4, i_6, \dots, i_m)$ are a permutation of the indices $(4, 6, \dots, m)$. These $(t-1)!$ cases are equivalent so

$$\begin{aligned} \Psi_{m+1} &= \epsilon \binom{1234 \dots m+1}{3124 \dots m+1} 2 (-2\pi)^{-t} R_{2121} R_{2323} \dots R_{m, m+1, m, m+1} + \dots \\ &= (-4\pi)^{-t} B_{m+1} + \dots \end{aligned}$$

where we have omitted those terms not involving B_{m+1} . Thus $c(B_{m+1}, \Psi_{m+1}) \neq 0$ so $\Psi_{m+1} \neq 0$.

Let $P = \sum_i c_i \Psi_m^i = 0$. Since C_m^1 involves fourth derivatives of the metric, $c(C_m^1, \Psi_m^i) = 0$ for $i > 1$. Since

$$R_{i_1 i_2 j_1 j_2 k k} = \frac{1}{2} (g_{i_1 j_1 / i_2 j_2 k k} + g_{i_2 j_2 / i_1 j_1 k k} - g_{i_1 j_2 / i_2 j_1 k k} - g_{i_2 j_1 / i_1 j_2 k k})$$

plus other terms which are quadratic on the 2-jets of the metric, we may replace $R_{i_1 i_2 j_1 j_2 k k}$ by this linear expression in computing $c(C_m^1, \theta_m^1)$. We may therefore restrict the sum defining θ_m^1 to $k=2$. By multiplying by 4^t we may further restrict the sum to collections (I, J) with $i_{2k-1} < i_{2k}$ and $j_{2k-1} < j_{2k}$ for $k \geq 1$. Since C_m^1 is divisible by $g_{11/2222}$, $i_1 = j_1 = 1$ and $i_2 = j_2 = 2$. By multiplying by $(t-1)!$ we may suppose $i_k = j_k = k$ for all k . Therefore

$$\begin{aligned} \theta_m^1 &= (-2\pi)^{-t} R_{1212, 22} R_{3434} \dots R_{m-1, m, m-1, m} + \dots \\ &= (-4\pi)^{-t} C_m^1 + \dots \end{aligned}$$

Therefore $c(C_m^1, P) = c_1 (-4\pi)^{-t} = 0$ which implies $c_1 = 0$.

If $m=2$, then $c_2=0$. Let $m>2$. If $i \geq 3$, θ_m^i is a polynomial in the 2-jets of the metric so $c(C_m^2, \theta_m^i) = 0$. Again, by multiplying by $(t-2)! 4^t$, we may restrict the sum defining θ_m^2 to $k=2$, $i_\nu = j_\nu = \nu$ for $\nu \geq 5$ and $i_1 < i_2$, $j_1 < j_2$, $i_3 < i_4$, $j_3 < j_4$. If $i_1 = 1$, then $I = J = (1, 2, \dots, m)$, while if $i_1 = 3$, then $I = J = (3, 4, 1, 2, 5, 6, \dots, m)$. Since these two cases are similar,

$$c(C_m^2, \theta_m^2) = 2 (-2\pi)^{-t} (2)^{-t} = 2 (-4\pi)^{-t} \neq 0 \text{ so } c_2 = 0.$$

If $m=2$, then $\theta_m^3 = (-8\pi)^{-1} \frac{1}{2} K^2$ where $K \neq 0$ is the scalar curvature. Thus $c_3 = 0$. If $m>2$, we evaluate $c(C_m^3, \theta_m^i)$ for $i \geq 3$. Since the index 2 appears 6 times in C_m^3 , we may restrict to $k=l=2$. This implies $c(C_m^3, \theta_m^4) = c(C_m^3, \theta_m^5)$

$= 0$. If $k = l = 2$ in Θ_m^6 , then the index '2' appears in *four* different variables so $c(C_m^3, \Theta_m^6) = 0$. By multiplying by $(t-1)!4^{t-1}$, we restrict the sum defining Θ_m^3 to range over $i_\nu = j_\nu = \nu$ for $\nu \geq 5$ and $i_3 < i_4$ and $j_3 < j_4$. This implies $i_3 = 2$ so (i_1, i_2, i_4) is a permutation of $(1, 3, 4)$. Consequently,

$$c(C_m^3, \Theta_m^3) = 6(-8\pi)^{-t} 4^{t-1} 2^{-t-1} = \frac{3}{4}(-4\pi)^{-t} \neq 0 \text{ so } c_3 = 0.$$

If $m = 2$, $c_4 = c_5 = c_6 = 0$. In computing $c(C_m^i, \Theta_m^v)$ for $i \geq 4$, the indices '1' and '2' each appear 4 times. As the roles of k and l are symmetrical in defining Θ_m^i , we multiply by 2 and set $k = 1$, $l = 2$. These indices appear in *four* different variables of Θ_m^6 so

$$c(C_m^4, \Theta_m^6) = c(C_m^5, \Theta_m^6) = 0.$$

Since the indices 1 and 2 appear in the same variable once in C_m^4 , but twice in every monomial of Θ_m^5 , $c(C_m^4, \Theta_m^5) = 0$. By multiplying by $t!4^t$, we restrict the sum defining Θ_m^4 to range over $i_\nu = j_\nu = \nu$. This shows

$$c(C_m^4, \Theta_m^4) = 2(-8\pi)^{-t} 4^t 2^{-t-1} = (-4\pi)^{-t} \neq 0 \text{ so } c_4 = 0.$$

Similarly, by multiplying by $(t-1)!4^t$, we may restrict the sum defining Θ_m^5 to range over $i_\nu = j_\nu = \nu$. This shows

$$c(C_m^5, \Theta_m^5) = 2(-8\pi)^{-t} 4^t 2^{-t-1} = (-4\pi)^{-t} \neq 0 \text{ so } c_5 = 0.$$

Finally, $c_6 = 0$ if $m \leq 4$. By multiplying by $2 \cdot 4^{t-1} (t-3)!$, we may restrict the sum defining Θ_m^6 to $k = 1$, $l = 2$, $i_\nu = j_\nu = \nu$ for $\nu \geq 7$, $j_2 < j_3$, $i_2 < i_3$, $i_5 < i_6$, and $j_5 < j_6$. Either $i_1 = 3$, $j_1 = 4$ or $i_1 = 4$, $j_1 = 3$. Similarly, either $i_4 = 5$, $j_4 = 6$ or $i_4 = 6$, $j_4 = 5$. By multiplying by 4 and combining all these cases, we conclude

$$\Theta_m^6 = (-2\pi)^{-t} 2 \varepsilon \begin{pmatrix} 3145267 \dots m \\ 4136257 \dots m \end{pmatrix} R_{1313} R_{1414} R_{2525} R_{2626} R_{7878}$$

$$R_{m-1, m, m-1, m} + \dots$$

so $c(C_m^6, \Theta_m^6) = (-4\pi)^{-t} \neq 0$ so $c_6 = 0$.

We now state and prove the generalisation of theorem 2.3 that we will use in §3 to complete the proof of theorem 1.3 :

Theorem 2.5

Let $m \geq 2$ be an even integer. Then :

- (a) $\mathcal{K}_n^m = 0$ if $n < m$.
- (b) $\mathcal{K}_m^m$ is one-dimensional and is spanned by E_m .
- (c) $\mathcal{K}_{m+2}^{m+1}$ is one-dimensional and is spanned by Ψ_{m+1} .
- (d) $\mathcal{K}_{m+2}^m$ is spanned by the $\{\Theta_m^i\}$ for $i = 1, \dots, 6$. It has dimension 6 if $m \geq 6$, it has dimension 5 if $m = 4$, and it has dimension 2 if $m = 2$.

Although (a) and (b) were proved earlier in [4], we review the proof here as we shall need some of the techniques involved.

If $A = g_{i_1 j_1 / a_1} \cdots g_{i_r j_r / a_r}$,

is a monomial of $\mathcal{P}^m$, let $L(A) = r$ be the length of A and let

$$\deg_k(A) = \delta_{i_1 k} + \delta_{j_1 k} + a_1(k) + \cdots + \delta_{i_r k} + \delta_{j_r k} + a_r(k)$$

be the number of times the index k appears in the monomial A . It is clear that $\text{ord}(A) = \sum_k \deg_k(A) - 2L(A)$.

Lemma 2.6

Let $P \in \mathcal{P}_n^m$ and suppose $c(A, P) \neq 0$. Then

(a) $\deg_l A$ is even for all l .

(b) If $P \in \mathcal{K}_n^m$, then $\deg_l(A) > 0$ for all l .

Proof: Fix an index l . Let $X = (x_1, \dots, x_m)$ be a coordinate system on M^m with

$$g_{ij}(x_0) = \delta_{ij} \quad \text{and} \quad g_{ij|k}(x_0) = 0.$$

Let $Y = (x_1, \dots, -x_l, \dots, x_m)$. Then $A(Y, G) = (-1)^{\deg_l A} A(X, G)$.

We can decompose

$$P(X, G) = \sum_A c(A, P) A(X, G).$$

Since P is invariant, $P(Y, G) = P(X, G)$ so $(-1)^{\deg_l A} = 1$ for all indices l and for all $c(A, P) \neq 0$. This proves (a).

We defined $r(P)(G_1) = P(G_1 \times 1)(x_1, \theta)$ where $G_1 \times 1$ is the product metric on $M_1 \times S^1$. The only additional relations imposed by considering such a metric are that

$$g_{im} \equiv \delta_{im} \quad \text{and} \quad g_{ij|a} \equiv 0 \quad \text{if} \quad a(m) \neq 0.$$

Therefore r can be defined on the generators of $\mathcal{P}^m$ by :

$$r(g_{ij|a}) = \begin{cases} 0 & \text{if } \deg_m(g_{ij|a}) \neq 0 \\ g_{ij|a} \in \mathcal{P}^{m-1} & \text{if } \deg_m(g_{ij|a}) = 0. \end{cases}$$

If $r(P) = 0$, then $\deg_m(A) = 0$ for every monomial A of P . Since P is invariant the form of P is unchanged by coordinate permutations so $\deg_l(A) \neq 0$ for all l .

If $P \in \mathcal{P}_n^m$ and if A is a monomial of P , $\sum_k \deg_k A = \text{ord}(A) + 2L(A)$ is even so $n = \text{ord}(A)$ is even. This proves

Cor. 2.7

$$\mathcal{P}_n^m = 0 \quad \text{if } n \text{ is odd.}$$

Before beginning the proof of theorem 2.5, we will need the following lemma which is concerned with invariance under the orthogonal group. If $\theta \in SO(1)$,

let $X = (\cos(\theta) x_1 + \sin(\theta) x_2, -\sin(\theta) x_1 + \cos(\theta) x_2, x_3, \dots, x_m)$.

$O(m)$ is generated by this action of $SO(2)$ and by the coordinate permutations. This action preserves the normalisation that

$$g_{ij}(X, G)(x_0) = \delta_{ij} \quad \text{and} \quad g_{ijk}(X, G)(x_0) = 0.$$

if $P \in \mathcal{P}^m$, P is invariant under this action of $SO(2)$ if and only if $P(X, G)(x_0) = P(X^\theta, G)(x_0)$ for all (X, G) and for all $\theta \in SO(2)$. Such invariance has the following algebraic consequences :

Lemma 2.8

Let $P \in \mathcal{P}^m$ be invariant under the action of $SO(2)$.

(a) Suppose $g_{12/\alpha}$ divides a monomial A of P . Then there is a monomial $A' \neq A$ of P divisible by $g_{11/\beta}$ for some β .

(b) Suppose $g_{ij/\alpha}$ divides a monomial A of P with $\alpha(2) > 0$. Then there is a monomial $A' = A$ of P divisible by $g_{kl/\beta}$ for some (k, l) with $\beta(1) = \alpha(1) + 1$ and $\beta(2) = \alpha(2) - 1$.

(c) The A' constructed in (a) and (b) is not arbitrary. It can be constructed from A by either (i) changing two '2' indices to '1' indices, (ii) changing two '1' indices to '2' indices, or (iii) by interchanging a '2' and a '1' index.

Of course, the use of '1' and '2' is for notational convenience only. This lemma holds for any pair of distinct indices and for any polynomial which is invariant under the appropriate action of $SO(1)$. Although this lemma was proved in [4], we give a brief sketch of the proof here as it is central to our discussion. We shall need a somewhat sharper version of this lemma in the proof of theorem 2.5 (d) which we will discuss later.

Decompose $P = \sum_\nu P_\nu$ into polynomials P_ν of total degree ν in the indices '1' and '2'. The P_ν are invariant separately under the action of $SO(2)$ so we may suppose that $P = P_\nu$ for some ν without loss of generality. For any $\theta \in SO(2)$,

$$\partial/\partial x_1^\theta = \cos(\theta) \partial/\partial x_1 + \sin(\theta) \partial/\partial x_2,$$

$$\partial/\partial x_2^\theta = \cos(\theta) \partial/\partial x_2 - \sin(\theta) \partial/\partial x_1.$$

The remainder of the coordinate frame is unchanged. We compute $g_{ij/k_1 \dots k_s}(X^\theta, G)$ by replacing each index '1' by ' $\cos(\theta) 1 + \sin(\theta) 2$ ' and each index '2' by ' $\cos(\theta) 2 - \sin(\theta) 1$ ' and expanding multilinearly. This decomposes $P(X^\theta, G)$ into

$$P(X^\theta, G) = \sum_i \cos(\theta)^{p-i} \sin(\theta)^i Q_i(X, G).$$

The terms with $i = 1$ come from changing one index '1' $\rightarrow$ '2' or '2' $\rightarrow$ '1' in the expansion.

Since P is $SO(2)$ invariant, we substitute $1 = \cos^2(\theta) + \sin^2(\theta)$ and $P(X, G) = P(X^\theta, G)$ to show $Q_{2p+1} = 0$ and $Q_{2i} = \binom{p}{i} P$. We use Q_1 in the proof of the lemma. Let $A = (g_{12/\alpha})^l A_0$ for $l > 0$. Then

$$A(X^\theta, G) = \cos(\theta)^p A(X, G) - l \cos(\theta)^{p-1} \sin(\theta) g_{11/\alpha} (g_{12/\alpha})^{l-1} A_0 + \dots$$

We have omitted those terms having a power of $\sin(\theta)$ other than 1 or which are not of the form $g_{11/\alpha} (g_{12/\alpha})^{l-1} A_0$. Since $Q_1 = 0$, there must be some other

monomial A' of P which is different from A and which cancels this contribution to Q_1 . Since the power of $\sin(\theta)$ is one, A' can be constructed from $g_{11|\alpha}(g_{12|\alpha})^{l-1}A_0$ by changing exactly one index. Consequently, A' can be constructed from A by changing two indices; we do not change any indices greater than 2. The possibilities are :

$$A = (g_{12|\alpha})^l A_0,$$

$$A'_1 = g_{11|\beta}(g_{12|\alpha})^{l-1} A_0 \text{ for } \beta(1) = \alpha(1) \pm 1 \text{ and } \beta(2) = \alpha(2) \mp 1,$$

$$A'_2 = g_{11|\alpha}(g_{12|\beta})(g_{12|\alpha})^{l-2} A_0 \text{ for } \beta(1) = \alpha(1) \pm 1 \text{ and } \beta(2) = \alpha(2) \mp 1,$$

$$A'_3 = g_{11|\alpha}(g_{12|\alpha})^{l-1} A_0,$$

where A'_0 is constructed from A_0 by changing exactly two indices less than 3. The first possibility is ruled out since $A \neq A'$, and any of the remaining monomials will satisfy the conditions of the lemma. The proof of (b) is similar and is therefore omitted.

We use lemma 2.8 to prove

Lemma 2.9

Let $P \in \mathfrak{K}_n^m$. If $c(A, P) \neq 0$, then $2L(A) \geq m$.

Proof: Let $A = g_{i_1 j_1 |\alpha_1} \dots g_{i_r j_r |\alpha_r}$ be a monomial of P . We will use lemma 2.8 to construct a new monomial A' so $\deg_k A' = 0$ for $k > 2L(A')$. Since applications of lemma 2.2 merely shuffle the indices around without changing the length, $\deg_k A' = 0$ for $k > 2L(A) = 2r$. Since $r(P) = 0$, $\deg_m A' \neq 0$ so $2r \geq m$ which will prove the lemma.

By performing a coordinate permutation, we may suppose $i_1 = 1$. By applying lemma 2.8 (a) if necessary, we may choose A so $j_1 = 1$. If i_2 and j_2 are distinct and greater than 1, we can apply lemma 2.8 (a) to assume $i_2 = j_2$ and make a coordinate permutation to assume $i_2 = j_2 = 2$. If $i_2 = 1$ and $j_2 \neq 1$, we may make a coordinate permutation to assume $j_2 = 2$. Thus we may choose A' so i_1, i_2, j_1, j_2 are less than 3. We proceed inductively to choose A' so the collection of indices $\{i_1, j_1, \dots, i_r, j_r\}$ range between 1 and r .

We now apply lemma 2.8 (b) to choose A' so $\alpha(k) = 0$ for $k > r + 1$ and continue inductively to choose A' so A' does not involve any indices greater than $2r$. This completes the proof.

We have now derived the necessary lemmas to begin the proof of theorem 2.5. Let $P \in \mathfrak{K}_n^m$. Since $|a_i| \geq 2$, $n = |a_1| + \dots + |a_r| \geq 2r \geq m$ by lemma 2.9. Therefore $P = 0$ if $n < m$ which proves (a). If $n = m$, m must be even and all these inequalities are equalities. Therefore $|a_i| = 2$ and $2r = m$. Since $\deg_k(A) \neq 0$ is even,

$$2m \leq \sum_k \deg_k A = 2r + n = 2m,$$

so $\deg_k A = 2$ for each index k and every monomial A of P .

If $P \in \mathcal{K}_m^m$, P is a polynomial in the 2-jets of the metric. We say that an index i touches itself in a monomial A of P if A is divisible by $g_{ii/kl}$ or by $g_{kl/ii}$ for some (k, l) . Choose a monomial A of P so the number of indices which touch themselves in A is maximal. Suppose the index i does not touch itself in A , then since i appears twice in A , i touches an index j which also does not touch itself. By applying lemma 2.8, we can rearrange the indices i and j to create a new monomial A' of P in which i touches itself. This contradicts the maximality in the choice of A and consequently every index touches itself in A . Since each index appears twice in A , A has the form

$$g_{i_1 i_1 / j_1 j_1} \cdots g_{i_t i_t / j_t j_t},$$

for $2t = m$ and for $(i_1, j_1, \dots, i_t, j_t)$ a permutation of the indices from 1 to m . By making a coordinate permutation, we can assume $\bar{A} = g_{11/22} \cdots g_{m-1, m-1/mm}$ is a monomial of P . Thus $P \neq 0$, $P \in \mathcal{K}_m^m$ implies $c(\bar{A}, P) \neq 0$. This implies $\mathcal{K}_m^m$ is at most one-dimensional. Since $E_m \neq 0$, $E_m \in \mathcal{K}_m^m$, this completes the proof of theorem 2.5 (b). This completes the proof of theorem 1.2 and provides a heat equation proof of the Chern-Gauss-Bonnet theorem.

Before beginning the proof of theorem 2.5 (c), we must prove the following lemma :

Lemma 2.10

Let $P \in \mathcal{K}_m^m$ be a polynomial in the 2-jets of the metric.

(a) If $c(A, P) \neq 0$, $g_{11/11}$, $g_{12/11}$, and $g_{11/12}$ do not divide A .

(b) If $A = g_{11/22} A_0$, $A_1 = g_{12/12} A_0$, and $A_2 = g_{22/11} A_0$, then

$$c(A, P) = c(A_2, P) = -\frac{1}{2} c(A_1, P).$$

(c) If $A = g_{11/23} A_0$, $A_1 = g_{12/13} A_0$, and $A_2 = g_{23/11} A_0$, then

$$c(A, P) = c(A_2, P) = -c(A_1, P).$$

Proof : We proved lemma 2.10 in [4] by using nonlinear changes of coordinates. We give a proof here using geodesic coordinates in the interests of brevity. The curvature tensor is given by

$$R_{ijkl} = \frac{1}{2} (g_{ik/jl} + g_{jl/ik} - g_{il/jk} - g_{jk/il}),$$

since we have assumed $g_{ij}(x_0) = \delta_{ij}$ and $g_{ij/k}(x_0) = 0$. In geodesic polar coordinates, the second derivatives of the metric can be expressed in terms of the R_{ijkl} 's. Since P is invariant and the R_{ijkl} 's are tensorial, this implies $P = P(R_{ijkl}(X, G))$ for any coordinate system. Since $g_{11/11}$, $g_{11/12}$, and $g_{12/11}$ do not appear in any of the R_{ijkl} polynomials, these variables divide no monomial of P . This proves (a). The variables $\{g_{11/22}, g_{22/11}, g_{12/12}\}$ appear only in

$$R_{1212} = R_{2121} = -R_{1221} = -R_{2112} = \frac{1}{2} (g_{1'122} + g_{22/11} - 2g_{12/12})$$

which proves (b). Finally, the variables $\{g_{11/23}, g_{23/11}, g_{12/13}, g_{13/12}\}$ appear only in $R_{1213} = R_{1312} = \text{etc.}$ Since

$$R_{1213} = \frac{1}{2} (g_{11/23} + g_{23/11} - g_{12/13} - g_{13/12}),$$

this proves (c).

Let $P \in \mathcal{U}_{m+2}^{m+1}$ with $P \neq 0$. By Cor. 2.7, $m+2$ is even so m is even. Let A be a monomial of P . By lemma 2.9, $2r = 2L(A) \geq m+1$. Since $m+1$ is odd, $2L(A) \geq m+2$. We compute :

$$m+2 = \text{ord}(A) = \sum_{\nu=1}^r |a_{\nu}| \geq 2r = 2L(A) \geq m+2.$$

This shows $m+2 = 2r$ and that $|a_{\nu}| = 2$ for all ν ; P is a polynomial in the 2-jets of the metric. Furthermore,

$$\sum_k \deg_k(A) = n + 2L(A) = 2(m+2) = 2(m+1) + 2.$$

Since $\deg_k(A) \geq 2$ is even for $k = 1, \dots, m+1$, one index appears 4 times in A and the remaining indices each appear twice.

We shall say that the index i touches the index j in A if $g_{ij/kl}$ or $g_{kl/ij}$ divides 4 for some (k, l) . Choose a monomial A of P so the number of indices which touch themselves in A is maximal. Let k be an index which appears twice in A and suppose k touches l where $k \neq l$ and where l is not the index which appears 4 times in A . Then as l also appears twice, neither k nor l touch themselves in A . By applying lemma 2.8, we can construct a new monomial A' of P in which k touches itself by altering only (k, l) indices. Since one more index touches itself in A' , this contradicts the maximality of A . By making a coordinate permutation if necessary, we may suppose that the unique index which appears 4 times in A is '2'. Then any index other than '2' appears twice and either touches itself or only touches the index '2'.

Next, suppose that the index '2' does not touch itself. It touches indices which appear twice and which therefore do not touch themselves. By applying lemma 2.8, we may construct a new monomial A' of P in which '2' touches itself. This contradicts the maximality of A . Either $A = g_{ij/22} A_0$ or $A = g_{22/ij} A_0$; by applying lemma 2.10 if necessary, we may assume the former without loss of generality. Neither i nor j is 2 by lemma 2.10. Thus i and j appear only twice in A so $i = j$. By making a coordinate permutation, we may suppose $i = j = 1$.

The index '2' appears exactly twice in the monomial A_0 . Suppose that '2' does not touch itself in A_0 . If $A_0 = g_{i2/j2} A'_0$, we may apply lemma 2.10 to show $g_{11/22} g_{ij/22} A'_0$ is also a monomial of P which contradicts the maximality of A . Thus we may express $A_0 = g_{ij/2k} A'_0$ or $A_0 = g_{2k/ij} A'_0$ where i, j, k are all distinct from 2. This shows $i = j$ so we may apply lemma 2.10 to assume $A_0 = g_{ii/2k} A'_0$ without loss of generality. The index '2' appears once in A'_0 and we apply a similar chain of reasoning to show we can choose $A = g_{11/22} g_{ii/2k} g_{jj/2l} A'_0$. Now k cannot touch another index which appears twice and therefore $k = l$. We may suppose that $k = l = 3$ for notational convenience. The indices $\{1, 2, i, j, 3\}$ are all distinct and do not appear in A'_0 .

We make a coordinate rotation in the '2' and '3' axes and compute the coefficient of $\cos^5(\theta) \sin(\theta) g_{11/22} g_{ii/22} g_{jj/23} A'_0$. Let $\bar{B}$ be a monomial which makes a non-zero contribution to this coefficient. Then $\bar{B}$ must transform to $g_{11/22} g_{ii/22} g_{jj/23} A'_0$ by changing exactly one '3' index to a '2' index or a '2' index to a '3' index. If $\bar{B}$ transforms by changing a '2' index to a '3' index, then $\deg_3 \bar{B} = 0$ which is impossible. Consequently, the 3 choices are :

$$A = g_{11/22} g_{44/23} g_{11/23} A_0, \quad A_2 = g_{11/23} g_{44/22} g_{11/23} A_0,$$

$$A_3 = g_{11/22} g_{44/22} g_{11/33} A_0.$$

A_3 is not a monomial of P since this would contradict the maximality in the choice of A . Therefore

$$P(X^\theta, G) = (c(A, P) + c(A_2, P)) \cos(\theta)^5 \sin(\theta) g_{11/22} g_{44/22} g_{11/23} A_0 +$$

other terms. Since P is $SO(2)$ invariant, $c(A, P) + c(A_2, P) = 0$. Since the indices '1' and i are distinct, we make a coordinate permutation to show $c(A, P) = c(A_2, P)$. Thus $c(A, P) = 0$ which is impossible. This shows that the index '2' touches itself in A_0 .

By making a co-ordinate permutation and applying lemma 2.10 if necessary, we may suppose $g_{11/22} g_{33/22} A'_0$ is a monomial of P . Each index which appears in A'_0 appears twice and does not touch the index '2' and therefore touches itself. By making a coordinate permutation, we may assume A has the form :

$$A = g_{11/22} g_{33/22} g_{44/55} \dots g_{mm/m+1, m+1} = B_{m+1}.$$

We summarise :

if $P \neq 0 \in \mathfrak{K}_{m+2}^{m+1}$, then $c(B_{m+1}, P) \neq 0$.

This shows $\mathfrak{K}_{m+2}^{m+1}$ is one-dimensional and proves theorem 2.5 (c).

In the proof of lemma 2.4, we showed that the matrix $c(C_m^i, \Theta_m^j)$ was invertible. We finish the proof of theorem 2.5 (d) by showing the linear functionals $c(C_m^i, P)$ separate $\mathfrak{K}_{m+2}^m$. The proof is somewhat technical. It uses the same techniques as those developed earlier and may be skipped without loss of continuity by the reader.

Let $P \neq 0 \in \mathfrak{K}_{m+2}^m$. Suppose $c(C_m^i, P) = 0$ for $1 \leq i \leq 6$. Let $c(A, P) \neq 0$ and decompose

$$A = g_{i_1 j_1 / a_1} \dots g_{i_r j_r / a_r} \text{ so that } 2r \geq m \text{ and } |a_1| \geq \dots \geq |a_r|.$$

Because $m+2 = \sum_i |a_i| \geq 2r \geq m$, either $2r = m$ or $2r = m+2$. Suppose that $2r = m$. Then

$$|a_2| \leq |a_1| = m+2 - |a_2| - \sum_{k>2} |a_k| \leq m+2 - |a_2| - 2(r-2) \\ = 6 - |a_2|.$$

If $|a_2| = 3$, then these inequalities are equalities so $|a_1| = 3$ and $|a_k| = 2$ for $k > 2$. If $|a_2| = 2$, then $|a_k| = 2$ for $k > 1$ so $|a_1| = m+2 - 2(r-1) = 4$. In either case, if $2r = m$ then

$$2m+2 = 2r + m + 2 = \sum_k \deg_k(A) \geq 2m.$$

Since $\deg_k(A) \geq 2$ is even, one index appears four times in A and the remaining indices each appear twice.

We consider first the case that $|a_1| = 4$ for some monomial A of P . By applying lemma 2.8, we may suppose that A has the form

$$A = g_{11/2222} A_0.$$

The index '2' appears nowhere else in A . Every other index appears exactly twice. Choose A of this form so the number of indices which touch themselves in A is maximal. Then every index distinct from 2 touches itself in A and appears twice. By making a coordinate permutation, we may assume that

$$A = g_{11/2222} g_{33/44} \cdots g_{m-1, m-1/mm} = C_m^1.$$

This is impossible since we assumed $c(C_m^1, P) = 0$.

Next, we suppose that $|a_1| = |a_2| = 3$. We apply lemma 2.8 to choose $c(A, P) \neq 0$ and

$$A = g_{ij/222} g_{kl/abc} A_0.$$

If the indices (c, d, e) are all distinct from '2', we may apply lemma 2.8 to choose a monomial A of P of the form :

$$A = g_{ij/222} g_{kl/aaa} A_0.$$

Since only one index can appear four times and the others all appear twice, this is impossible. We may therefore assume that A has the form :

$$A = g_{ij/222} g_{kl/2bc} A'_0.$$

Since the index '2' appears exactly 4 times in A , b and c are distinct from '2' so we may apply lemma 2.8 to choose A of the form :

$$A = g_{ij/222} g_{kl/244} A'_0.$$

The indices '2' and '4' appear nowhere else in A . Choose A of this form so the number of indices which touch themselves in A is maximal. Then every index other than '2' and '4' touches itself. By making a coordinate permutation we may assume A has the form :

$$A = g_{11/222} g_{33/244} g_{55/66} \cdots g_{m-1, m-1/mm} = C_m^2.$$

This is impossible since we assumed $c(C_m^2, P) = 0$.

Therefore, if A is a monomial of P , $2L(A) = 2r = m + 2$. Consequently, $m + 2 = \sum |a_k| \geq 2r \geq m + 2$ so each $|a_k| = 2$. This shows that P is a polynomial in the 2-jets of the metric alone so we can apply lemma 2.10 to the monomials of P . Since $\sum_k \deg_k A = 2r + \text{ord}(A) = 2m + 4$, and since $\deg_k(A) \geq 2$ is even for all k , either one index appears six times and the remaining indices each appear twice or two indices each appear four times and the remaining indices each appear twice.

Before considering these two cases further, we must prove a somewhat sharper version of lemma 2.8.

Lemma 2.11

Let $0 \neq P \in \mathcal{K}_n^m$ and let $c(A, P) \neq 0$. Suppose that $\deg_1(A) = 2$ and that A has the form $A = g_{1i/jk} A_0$ (or $g_{jk/1i} A_0$). Then there is a monomial

A' of P of the form $A' = g_{2i|j'k'} A'_0$ (respectively $g_{j'k'|2i} A'_0$) such that $c(A, P) c(A', P) < 0$. A' is constructed from A by interchanging a '1' and a '2' index. If j and k are distinct from '1' and '2', then $j' = j$ and $k' = k$. In this case, $A' = g_{2i|jk} A'_0$ (respectively $g_{jk|2i} A'_0$) where A'_0 is constructed from A_0 by changing a '2' index to a '1' index.

Proof: We argue as in the proof of lemma 2.8. We make a coordinate rotation in the '1' and '2' axes. We suppose A has the form $A = g_{1i|jk} A_0$ as the other case is proved similarly. Let $v = \deg_1 A + \deg_2 A = 2 + \deg_2 A$ and let μ be the multiplicity with which $g_{1i|jk}$ divides A . Let $\rho = 2$ if $i = 1$ and let $\rho = 1$ otherwise. Then

$$A(X^\theta, G) = \cos^v \theta A(X, G) + \mu \rho \cos^{v-1} \theta \sin \theta g_{2i|jk} A_0 + \dots$$

Since $Q_1 = 0$ in the expansion of $P(X^\theta, G)$, there must be some monomial A' of P which helps to cancel this contribution to Q_1 . A' can be transformed into $g_{2i|jk} A_0$ by changing a single '1' $\rightarrow$ '2' or '2' $\rightarrow$ '1'. Since $\deg_1 A' \geq 2$ and $\deg_1(g_{2i|jk} A_0) = 1$, A' can be transformed into $g_{2i|jk} A_0$ by changing a single '1' $\rightarrow$ '2'. Therefore

$$A'(X^\theta, G) = \cos^v \theta A'(X, G) + c \cdot \cos^{v-1} \theta \sin \theta g_{2i|jk} A_0 + \dots$$

where $c > 0$. Since A' helps cancel the contribution of $g_{2i|jk} A_0$ made by A to Q_1 , $c(A, P)$ and $c(A', P)$ have opposite signs. A' is constructed from A by interchanging a '2' and a '1' index. If we express $A' = g_{2i'|j'k'} A'_0$ and if $i > 3$, then $i = i'$ since we are only changing '1' and '2' indices. If $i = 1$, then $i' = 1$ since we are interchanging a '2' and a '1' index. If $i = 2$ and if $i' \neq 1$, then $i' = 1$ so $A' = A$. This is impossible because $c(A, P) c(A', P) < 0$. Therefore $i = i'$ in all cases which completes the proof.

Suppose some index appears six times in some monomial A of P . By making a coordinate permutation, we may suppose that $\deg_2 A = 6$ and that $\deg_k A = 2$ for $k \neq 2$. Let A be chosen in this form so the number of indices which touch themselves in A is maximal. We use lemma 2.8 to show that if $k \neq 2$, either k touches itself in A or it touches only the index '2' in A . Suppose A has the form $A = g_{2i|2j} A_0$ for some (i, j) , then $A' = g_{ii|22} A_0$ is also a monomial of P by lemma 2.10. The index i does not touch the index j in A and any index k which touches itself in A also touches itself in A' . Therefore A' is also maximal with respect to the number of indices which touch themselves. By lemma 2.10, $i \neq 2$ and $j \neq 2$. Since i touches an index distinct from j , it touches itself in A' so $i = j$. This implies the number of indices which touch themselves in A' is greater than the number of indices which touch themselves in A which is impossible. Therefore $g_{2i|2j}$ does not divide A for any (i, j) .

The only variables involving the index '2' which can divide A have the form $g_{ii|22}$, $g_{ii|2j}$, $g_{2j|ii}$, or $g_{2j|ii}$ where i and j are distinct from '2'. By using lemma 2.10, we can construct A so the only variables involving the index '2' which divide A have the form $g_{ii|22}$ or $g_{ii|2j}$. Suppose $A = g_{ii|2j} A_0$. We suppose $j = 1$ for notational convenience. By applying lemma 2.11 to the indices '1' and '2' and to the variable $g_{ii|12}$, we can construct $A' = g_{ii|22} A'_0$ by changing a single index '2' to '1' in A_0 ; $\deg_1 A' = 2$ and $\deg_2 A' = 6$. Since the only variables

which divide A involving the index '2' are of the form $g_{aa/22}$ or $g_{aa/2b}$, there are only two cases :

$$(1) \quad A = g_{ii/12} g_{aa/2b} A_1, \quad A' = g_{ii/22} g_{aa/1b} A_1 \quad (b \neq 2),$$

$$(2) \quad A = g_{ii/12} g_{aa/22} A_1, \quad A' = g_{ii/22} g_{aa/12} A_1.$$

If A has the form (1), any index which touches itself in A touches itself in A' so A' is maximal as well. This implies $b = j$ so one more index touches itself in A' than does in A which is impossible. If A has the form (2), we may make a coordinate permutation interchanging the indices i and a to show $c(A', P) = c(A, P)$ which contradicts the fact $c(A, P) c(A', P) < 0$.

This shows that any variable which involves the index '2' dividing A has the form $g_{ii/22}$. Therefore any index $k \neq 2$ touches itself and appears exactly twice in A . By making a coordinate permutation, we may suppose

$$A = g_{11/22} g_{33/22} g_{44/22} g_{55/66} \cdots g_{m-1, m-1/mm} = C_m^3.$$

This contradicts the fact that C_m^3 is not a monomial of P by hypothesis.

Therefore, if A is a monomial of P , two indices each appear four times in A and the remaining indices each appear twice in A . We choose A so $\deg_1 A = \deg_2 A = 4$ and so $\deg_k A = 2$ for $k > 2$. We can decompose A uniquely as a product $A = A_1 A_2$ so that any variable which divides A_1 involves either the index '1' or the index '2' or both and so that none of the variables which divide A_2 involves either the index '1' or '2'. Let $f(A) = L(A_1)$ be the number of variables (counted with multiplicity) which involve the index '1' or '2' which divide A .

Lemma 2.12

There exists a monomial A of P divisible by $g_{11/22}$ so $f(A)$ is minimal.

Proof: Choose any monomial A of P with $f(A)$ minimal. We show first A can be chosen so $g_{11/jk}$ divides A for some (j, k) . Since the index '1' appears in A four times, either $g_{11/jk}$ or $g_{jk/11}$ divides A . Let $A = g_{11/jk} A_0$ (or $g_{11/jk} A_0$). If $i > 2$, we apply lemma 2.11 to the indices i and '1' and to the variable $g_{11/jk}$ (or $g_{jk/11}$) to construct $A' = g_{11/j'k'} A'_0$ (or $g_{j'k'/11} A'_0$). Since we are replacing a '1' index by a i index in A'_0 or (j, k) , the number of variables involving the index '1' or '2' dividing A is not increased so $f(A') \leq f(A)$. If $A' = g_{j'k'/11} A'_0$, then $A'' = g_{11/j'k'} A'_0$ is also a monomial of P .

This shows that we may choose A of the form $A = g_{11/jk} A_0$ for some (j, k) with $f(A)$ minimal. We apply lemma 2.11 to the indices (j, k) and '2' and to the variable $g_{11/jk}$ to construct $A' = g_{11/22} A_0$. Since $f(A') \leq f(A)$, $f(A)$ is minimal. This completes the proof of the lemma.

Let A be a monomial of P which is divisible by $g_{11/22}$ so $f(A)$ is minimal. Among all such A , choose A so the number of indices $k > 2$ which touch themselves in A is maximal. We may use lemma 2.10 to replace any variables $g_{a1/aj}$ by $g_{a1/ij}$ without reducing the number of indices which touch themselves in A and without changing $f(A)$. We may therefore assume A is not divisible by $g_{a1/aj}$ for any (i, j) . The arguments given previously show that any index $k > 2$ either touches itself in

A or only touches the indices '1' and '2'. Let $A = A_1 A_2 = g_{11/22} A_0 A_2$. Since $\deg_1 A_2 + \deg_2 A_2 = 0$, every index which appears in A_2 touches some index other than '1' or '2'. Therefore every index which appears in A_2 touches itself and does not appear in A_1 . By making a coordinate rotation, we can assume that $A_2 = g_{2i+1, 2i+1/2i+2, 2i+2} \cdots g_{m-1, m-1/m}$ for $f(A) = i$. Then $\deg_k(A_1) = 2$ for $k > 2$.

Since $\deg_1 A + \deg_2 A = 8$, $\deg_1 A_0 + \deg_2 A_0 = 4$. Each variable which divides A_0 involves either the index '1' or the index '2'. This shows $L(A_0) \leq 4$ so $f(A) \leq 5$. We complete the proof of theorem 2.5 by applying downward induction on $f(A)$. Suppose $f(A) = 5$. Because $L(A_0) = 4$, each variable which divides A_0 can only involve one index '1' or '2'. These variables must have the form $g_{aa/1b}$, $g_{aa/2b}$, $g_{1b/aa}$, or $g_{2b/aa}$. We can use lemma 2.10 to eliminate the first two possibilities. We may assume A_0 is a variable in the $\{g_{1b/aa} g_{2b/aa}\}$ variables for $2 < a, b$. There are exactly two distinct variables $x \neq y > 2$ which touch the indices '1' and '2'. If x touches both '1' and '2' in A , then y touches both '1' and '2'. Otherwise, we may choose the notation so x touches only the index '1' and so y touches only the index '2'. There are two possible forms for A up to coordinate permutation :

$$(1) \quad A = g_{11/22} g_{1x/33} g_{1x/44} g_{2y/55} g_{2y/66} A_2,$$

$$(2) \quad A = g_{11/22} g_{1x/33} g_{1y/44} g_{2x/55} g_{2y/66} A_2.$$

All the monomials of the first form have the same coefficient in P as do all the monomials of the second form. Suppose A has the first form. We apply lemma 2.11 to the indices x and '1' and the variable $g_{1x/33}$ to construct

$$A' = g_{11/22} g_{11/33} g_{xx/44} g_{2y/55} g_{2y/66} A_2,$$

$$\text{or} \quad A' = g_{1x/22} g_{11/33} g_{1x/44} g_{2y/55} g_{2y/66} A_2.$$

The first case would contradict the minimality of $f(A)$ since $f(A') = 4$. We apply lemma 2.11 to the indices x and '1' and to the variable $g_{1x/44}$ to construct

$$A'' = g_{1x/22} g_{1x/33} g_{11/44} g_{2y/55} g_{2y/66} A_2,$$

$$\text{or} \quad A'' = g_{xx/22} g_{11/33} g_{11/44} g_{2y/55} g_{2y/66} A_2 \text{ from } A'.$$

In the first case, $c(A'', P) = c(A', P)$ which contradicts the fact $c(A'', P) c(A', P) < 0$. We apply lemma 2.10 to the second case to show that

$$B = g_{22/xx} g_{11/33} g_{11/44} g_{2y/55} g_{2y/66} A_2,$$

is a monomial of P . We apply lemma 2.11 to the indices y and '2' and to the variable $g_{2y/55}$ to construct

$$B' = g_{2y/xx} g_{11/33} g_{11/44} g_{22/55} g_{2y/66} A_2,$$

$$\text{or} \quad B' = g_{22/xx} g_{11/33} g_{11/44} g_{22/55} g_{yy/66} A_2.$$

The first case is impossible since $c(B', P) = c(B, P)$. The second case is impossible since $f(B') = 4 < f(A)$. Therefore, no monomial of the form (1) appears in P .

Suppose A has the form (2). We apply lemma 2.11 to the indices x and y and to the variable $g_{1\sigma/33}$ to construct

$$A' = g_{11/22} g_{1y/33} g_{1y/44} g_{2\sigma/55} g_{2\sigma/66} A_2,$$

or
$$A' = g_{11/22} g_{1y/33} g_{1\sigma/44} g_{2\sigma/55} g_{2y/66} A_2.$$

The first case is impossible since this is of the form (1) and no monomial of this form appears in P . The second case is impossible since $c(A', P) = c(A, P)$. Therefore $f(A) \leq 4$.

Suppose that $f(A) = 4$; if $A = g_{11/22} A_0 A_2$, then $L(A_0) = 3$. Among the three variables which divide A_0 , one variable contains two indices '1' and '2' while the remaining two variables only contain one index '1' or '2' each. Using lemma 2.10, we can choose A in one of the following forms:

$$(1) \quad A = g_{11/22} g_{11/33} g_{2\sigma/44} g_{2\sigma/55} A_2,$$

$$(2) \quad A = g_{11/22} g_{12/33} g_{1\sigma/44} g_{2\sigma/55} A_2,$$

$$(3) \quad A = g_{11/22} g_{1\sigma/2y} g_{1\sigma/44} g_{2y/55} A_2,$$

$$(4) \quad A = g_{11/22} g_{1\sigma/2y} g_{1y/44} g_{2\sigma/55} A_2.$$

Suppose that A has the form (1). We apply lemma 2.11 to the indices x and '2' and to the variable $g_{2\sigma/44}$ to construct

$$A' = g_{11/22} g_{11/33} g_{22/44} g_{2\sigma/55} A_2,$$

or
$$A' = g_{11/2\sigma} g_{11/33} g_{22/44} g_{2\sigma/55} A_2.$$

The first case is impossible since $f(A') = 3 < f(A)$. We apply lemma 2.11 to the indices x and '2' and to the variable $g_{2\sigma/55}$ to construct

$$A'' = g_{11/2\sigma} g_{11/33} g_{2\sigma/44} g_{22/55} A_2,$$

or
$$A'' = g_{11/2\sigma} g_{11/33} g_{22/44} g_{22/55} A_2.$$

The first case is impossible since $c(A'', P) = c(A, P)$. In the second case, by making a coordinate permutation and by using lemma 2.10, we construct

$$B = g_{55/11} g_{66/11} g_{33/22} g_{44/22} g_{77/88} \cdots g_{m-1, m-1/mm} = C_m^6$$

so $c(B, P) \neq 0$. This contradicts the hypothesis that $c(C_m^6, P) = 0$.

Suppose that A has the form (2). We apply lemma 2.8 to the variable $g_{12/33}$ to construct A' divisible by $g_{11/33}$. Since $\deg_1 A' + \deg_2 A' = 8$ and since no

index appears six times in A' , $\deg_1 A' = \deg_2 A' = 4$. We can use lemma 2.10 to replace the variable $g_{12/12}$ if it appears in A' by $g_{11/22}$. This constructs A'' of the form

$$A'' = g_{11/22} g_{11/33} g_{2x/44} g_{2y/55} A_2,$$

which has already been shown not to be a monomial of P .

Suppose that A has the form (3). We apply lemma 2.11 to the indices x and '1' and to the variable $g_{x/2y}$ to construct

$$A' = g_{11/22} g_{11/2y} g_{xx/44} g_{2y/55} A_2,$$

or
$$A' = g_{1x/22} g_{11/2y} g_{1x/yy} g_{2y/55} A_2.$$

The first case is impossible since $f(A') = 3 < f(A)$. We apply lemma 2.11 to the indices x and '2' and to the variable $g_{1x/yy}$ to construct

$$A'' = g_{1x/22} g_{11/2y} g_{12/44} g_{xy/55} A_2,$$

or
$$A'' = g_{x/22} g_{11/2y} g_{12/44} g_{2y/55},$$

or
$$A'' = g_{1x/2x} g_{11/2y} g_{12/44} g_{2y/55} A_2.$$

The first of these three cases is impossible since $f(A') = 3 < f(A)$. In the second case, we apply lemma 2.11 to the indices x and y and to the variable $g_{11/yy}$ to construct $B = g_{1x/22} g_{11/yy} g_{12/44} g_{2x/55} A_2$ from A' . We apply lemma 2.11 to the indices x and '1' and to the variable $g_{1x/22}$ to construct

$$B' = g_{11/22} g_{11/yy} g_{2x/44} g_{2x/55} A_2,$$

or
$$B' = g_{11/22} g_{1x/yy} g_{12/44} g_{2x/55} A_2,$$

The first case is of form (1) and the second case is of form (2) which is impossible. Therefore

$$A'' = g_{1x/2x} g_{11/2y} g_{12/44} g_{2y/55} A_2,$$

is a monomial of P . We apply lemma 2.10 to construct

$$B = g_{12/xx} g_{11/2y} g_{12/44} g_{2y/55} A_2,$$

from A'' . We apply lemma 2.11 to the indices y and '2' and to the variable $g_{11/2y}$ to construct

$$B' = g_{1y/xx} g_{11/22} g_{12/44} g_{2y/55} A_2,$$

or
$$B' = g_{12/xx} g_{11/22} g_{1y/44} g_{2y/55} A_2,$$

or
$$B' = g_{12/xx} g_{11/22} g_{12/44} g_{yy/55} A_2.$$

The first two cases are in the form (2) and the last has $f(B') = 3 < f(A)$. This contradiction shows A does not have form (3).

Suppose A has form (4). We apply lemma 2.11 to the indices x and y and to the variable $g_{1x/2y}$ to construct

$$A' = g_{11/22} g_{1y/2x} g_{1y/44} g_{2x/55} A_2,$$

or
$$A' = g_{11/22} g_{1y/2y} g_{1x/44} g_{2x/55} A_2.$$

The first case is of form (3) which is impossible. We apply lemma 2.10 to A' to construct

$$A'' = g_{11/22} g_{12/4y} g_{1x/44} g_{2x/55} A_2.$$

This is of form (2) which is impossible.

This shows $f(A) \leq 3$. If $f(A) = 3$, then $L(A_0) = 2$. We analyse the possible forms for A in this case.

Lemma 2.13. Let $\bar{A}$ be a monomial of P with $f(\bar{A}) = 3$ minimal. Then there is some monomial A of P with $f(A) = 3$ in one of the following forms:

- | | |
|--|--|
| (1) $A = g_{11/22} g_{11/33} g_{22/44} A_2,$ | (2) $A = g_{11/22} g_{11/2x} g_{33/2x} A_2,$ |
| (3) $A = g_{11/22} g_{12/33} g_{12/44} A_2,$ | (4) $A = g_{11/22} g_{1x/2y} g_{1x/2y} A_2,$ |
| (5) $A = g_{11/22} g_{1x/2y} g_{1y/2x} A_2,$ | (6) $A = g_{11/22} g_{1x/2y} g_{2x/1y} A_2,$ |
| (7) $A = g_{11/22} g_{1x/2y} g_{2y/1x} A_2.$ | |

Proof: Express

$$A = g_{11/22} g_{ab/cd} g_{ij/kl} A_2 \text{ for } \deg_1 A_2 + \deg_2 A_2 = 0.$$

Suppose

$$\deg_1(g_{ab/cd}) + \deg_2(g_{ab/cd}) = 3.$$

By making a coordinate permutation and using lemma 2.10 to replace $g_{22/11}$ by $g_{11/22}$ if necessary, we may suppose $\deg_1(g_{ab/cd}) = 2$. Since the index '1' appears twice in $g_{ab/cd}$, we may use lemma 2.10 to assume $g_{ab/cd} = g_{11/2x}$. Since any index other than x touches itself, we can use lemma 2.10 to suppose

$$A = g_{11/22} g_{11/2x} g_{33/2x} A_2,$$

which is of form (2). If

$$\deg_1(g_{ab/cd}) + \deg_2(g_{ab/cd}) = 1, \text{ then } \deg_1(g_{ij/kl}) + \deg_2(g_{ij/kl}) = 3$$

which is similar.

We therefore suppose $\deg_1(g_{ab/ed}) + \deg_2(g_{ab/ed}) = 2$. Suppose $\deg_1(g_{ab/ed}) = 2$ and $\deg_2(g_{ij/kl}) = 2$. Since A can be chosen so $g_{ai/aj}$ does not divide A for any (a, i, j) , '1' touches itself in A_0 as does '2'. By applying lemma 2.10, we may choose $A = g_{11/22} g_{11/33} g_{22/44} A_2$ which is of form (1). We therefore suppose that

$$\deg_1(g_{ab/ed}) = \deg_2(g_{ab/ed}) = \deg_1(g_{ij/kl}) = \deg_2(g_{ij/kl}) = 1.$$

If the index '1' touches '2' in A_0 , we may take

$$g_{ab/ed} = g_{12/33} \quad \text{so} \quad A = g_{11/22} g_{12/33} g_{12/44} A_2,$$

is of form (3). Finally, if the index '1' does not touch the index '2' in A , A_0 must be a monomial in the $\{g_{1a/2b}, g_{2a/1b}\}$ variables for $a \neq b$. This gives rise to 4 cases which are listed under (4)–(7).

We use lemma 2.13 to show that $f(A) = 3$ is impossible. If A has form (1) then

$$A' = g_{11/22} g_{44/11} g_{33/22} g_{55/66} \cdots g_{m-1, m-1/mm} = C_m^4,$$

is also a monomial of P which contradicts the hypothesis that $c(C_m^4, P) = 0$. If A has form (2), we apply lemma 2.11 to the indices x and '2' and to the variable $g_{11/22}$ to construct

$$A' = g_{11/22} g_{11/22} g_{33/xx} A_2,$$

$$\text{or} \quad A' = g_{11/2x} g_{11/22} g_{33/2x} A_2.$$

The first case is impossible since $f(A') = 2 < f(A)$. The second case is impossible since $c(A', P) c(A, P) < 0$ but $A' = A$. If A has form (3), we apply lemma 2.8 to the indices '1' and '2' and to the variable $g_{12/33}$ to construct $A' = g_{11/33} \cdots$. Since no monomial of P contains an index which appears 6 times, $\deg_1 A' = \deg_2 A' = 4$. We apply lemma 2.10 if necessary to construct

$$A'' = g_{11/22} g_{11/33} g_{22/44} A_2.$$

This is of form (1) which is impossible.

If A has form (4), we apply lemma 2.11 to the indices x and '1' and to the variable $g_{1x/2y}$ to construct

$$A' = g_{11/22} g_{11/2y} g_{xx/2y} A_2,$$

$$\text{or} \quad A' = g_{1x/22} g_{11/2y} g_{1x/2y} A_2.$$

The first case is of form (2) which is impossible. We apply lemma 2.11 to the indices x and '1' and to the variable $g_{1x/2y}$ to construct

$$A'' = g_{1x/22} g_{1x/2y} g_{11/2y} A_2,$$

or
$$A'' = g_{xx/22} g_{11/2y} g_{11/2y} A_2.$$

The first case is impossible since $A = A'$. We apply lemma 2.11 to the indices y and '2' and to the variable $g_{11/2y}$ to construct

$$A''' = g_{xx/2y} g_{11/22} g_{11/2y} A_2,$$

or
$$A''' = g_{xx/22} g_{11/22} g_{11/y y} A_2.$$

The first case is of form (2) which is impossible. We apply lemma 2.10 to construct

$$B = g_{22/xx} g_{11/22} g_{11/y y} A_2,$$

which is of form (1). This contradiction shows A is not of form (4).

Suppose A is of form (5). We apply lemma 2.11 to the indices x and y and to the variable $g_{1x/2y}$ to construct

$$A' = g_{11/22} g_{1y/2x} g_{1y/2x} A_2,$$

or
$$A' = g_{11/22} g_{1y/2y} g_{1x/2x} A_2.$$

The first case is of form (4) which is impossible. We apply lemma 2.10 to construct

$$A'' = g_{11/22} g_{12/y y} g_{12/x x} A_2,$$

which is of form (3) which is impossible.

Suppose A is of form (6). We apply lemma 2.11 to the indices x and '2' and to the variable $g_{1x/2y}$ to construct

$$A' = g_{11/22} g_{12/2y} g_{xx/1y} A_2,$$

or
$$A' = g_{11/2x} g_{12/2y} g_{2x/1y} A_2.$$

In the first case, we apply lemma 2.10 to construct

$$A'' = g_{11/22} g_{22/1y} g_{xx/1y} A_2,$$

which is of form (2) which is impossible. We apply lemma 2.10 to the second case to construct

$$A'' = g_{11/2x} g_{22/1y} g_{2x/1y} A_2.$$

We apply lemma 2.11 to the indices x and '2' and to the variable $g_{11/2x}$ to construct

$$B = g_{11/22} g_{2x/1y} g_{2x/1y} A_2,$$

or
$$B = g_{11/22} g_{22/1y} g_{22/1y}.$$

We apply lemma 2.10 to the first case to construct

$$B' = g_{22/11} g_{22/1y} g_{22/1y} A_2$$

which is of form (4) after interchanging the indices '1' and '2'. We apply lemma 2.10 to the second case to construct

$$B' = g_{22/11} g_{22/1y} g_{22/1y},$$

which is of form (2) after interchanging the indices '1' and '2'. This is impossible.

Suppose A is of form (7). We apply lemma 2.11 to the indices x and y and to the variable $g_{12/2y}$ to construct

$$A' = g_{11/22} g_{1y/22} g_{2y/12} A_2,$$

or
$$A' = g_{11/22} g_{1y/2y} g_{22/12} A_2.$$

The first case is of form (6) which is impossible. We apply lemma 2.10 to the second case to construct

$$A'' = g_{11/22} g_{12/yy} g_{12/xx} A_2.$$

This is of form (3) which is impossible.

This contradiction shows $f(A) \leq 2$. Since

$$\deg_1 A + \deg_2 A = 8, \quad A = g_{11/22} g_{ij/kl} A_2,$$

where the variable $g_{ij/kl}$ involves only the indices '1' and '2'. We may therefore apply lemma 2.10 to construct

$$A' = g_{11/22} g_{11/22} g_{33/44} \cdots g_{m-1, m-1/mm} = C_m^5$$

as a monomial of P . This contradicts the assumption that $c(C_m^5, P) = 0$.

We have exhausted all the possible cases. Therefore, $c(C_m^i, P) = 0$ for $1 \leq i \leq 6$ implies that $P = 0$ and this completes the proof of theorem 2.5.

3. Evaluation of normalising constants

In this section, we will complete the proof of theorem 1.3. Fix $2n = m + 2$. Since $a_n^m \in \mathcal{K}_{m+2}^m$, we may use theorem 2.5 to decompose $a_n^m = \sum_i c(i, m) \Theta_m^i$.

Lemma 3.1

We can choose the $c(i, m)$ so $c(i, m) = c_i$ is independent of the dimension m . This decomposes $a_n^m = \sum_i c_i \Theta_m^i$. [If $m \geq 6$, the Θ_m^i are linearly independent so the $c(i, m)$ are uniquely determined by this equation. If m is 2 or 4,

however, the Θ_m^i are linearly dependent so there is a certain degree of freedom in the choice of $c(i, 2)$ and $c(i, 4)$. We eliminate this indeterminacy by requiring $c(i, m) = c(i)$ is independent of m .]

Proof: Let M_i be Riemannian manifolds of even dimensions m_i with metrics G_i for $i = 1, 2$. Let $M = M_1 \times M_2$ with the product metric $G = G_1 \times G_2$ let $m = m_1 + m_2$, and let $x = (x_1, x_2)$. The normalising constants in the definition of the Θ_m^i were chosen so that :

$$\Theta_m^i(G)(x) = \Theta_{m_1}^i(G_1)(x_1) E_{m_2}(G_2)(x_2) + E_{m_1}(G_1)(x_1) \Theta_{m_2}^i(G_2)(x_2).$$

By lemma 2.2 :

$$a_n^m(G)(x) = \sum_{2p+2q=m+2} a_p^{m_1}(G_1)(x_1) a_q^{m_2}(G_2)(x_2).$$

By theorem 1.2 (a),

$$a_p^{m_1}(G_1)(x_1) = 0 \text{ for } 2p < m_1 \text{ and } a_q^{m_2}(G_2)(x_2) = 0 \text{ for } 2q < m_2.$$

Therefore, the only non-zero terms in this sum arise when either $2p = m_1$ and $2q = m_2 + 2$ or $2p = m_1 + 2$ and $2q = m_2$. By theorem 1.2 (b),

$$a_p^{m_1}(G_1)(x_1) = E_{m_1}(G_1)(x_1) \text{ and } a_q^{m_2}(G_2)(x_2) = E_{m_2}(G_2)(x_2),$$

$$\text{if } 2p = m_1 \text{ and } 2q = m_2.$$

Therefore

$$a_n^m(G)(x) = a_{n_1}^{m_1}(G_1)(x_1) E_{m_2}(G_2)(x_2) + E_{m_1}(G_1)(x_1) a_{n_2}^{m_2}(G_2)(x_2),$$

for $2n_1 = m_1 + 2$ and $2n_2 = m_2 + 2$. This implies that :

$$\begin{aligned} \sum_i c(i, m) \Theta_m^i(G)(x) &= \sum_i c(i, m) \{ \Theta_{m_1}^i(G_1)(x_1) E_{m_2}(G_2)(x_2) + E_{m_1}(G_1)(x_1) \Theta_{m_2}^i(G_2)(x_2) \} \\ &= \sum_i c(i, m_1) \Theta_{m_1}^i(G_1)(x_1) E_{m_2}(G_2)(x_2) \\ &\quad + \sum_i c(i, m_2) E_{m_1}(G_1)(x_1) \Theta_{m_2}^i(G_2)(x_2). \end{aligned}$$

If $m_1 \geq 6$, the $c(i, m)$ and $c(i, m_1)$ are uniquely determined so this equation implies $c(i, m) = c(i, m_1)$. If $m_1 \leq 4$, we can define $c(i, m_1) = c(i, 6)$ as this stable value.

We show $c_1 = 1/6$ and $c_2 = 1/12$ by applying the formulas developed in [5] to the case $m = 4$. Let $0 \leq p \leq 4$ and let ∇^p be the Levi-Civita connection on $C^\infty(\Lambda^p T^*M)$. Let X be a normalised coordinate system at x_0 ;

$$g_{ij}(X, G)(x_0) = \delta_{ij} \text{ and } g_{ij/k}(X, G)(x_0) = 0.$$

Let $e = (e_1, e_2, e_3, e_4)$ be a local orthonormal frame for TM which agrees with the coordinate frame at x_0 . Let W_{ij}^p be the curvature of the connection ∇^p ;

$$W_{ij}^p(\omega) = \{ \nabla_{e_i}^p \nabla_{e_j}^p - \nabla_{e_j}^p \nabla_{e_i}^p - \nabla_{[e_i, e_j]}^p \} \omega \text{ for any } \omega \in \Lambda^p T^*M.$$

The local frame e induces local frames on $\Lambda^p T^*M$. Relative to these induced local frames, we can regard $W_{ij}^p(x_0)$ as a matrix whose entries are linear in the second derivatives of the metric. If the variable $g_{ab/cd}$ is an entry of this matrix then both the indices i and j appear at least once in the variable $g_{ab/cd}$.

The Bochner or reduced Laplacian is given at x_0 by the formula :

$$D_p^4 \nabla = - \sum_k \partial^2 / \partial x_k^2.$$

This operator and the ordinary Laplacian have the same second and first order symbols. We define

$$\mathcal{E}_p = D_p^4 - D_p^4 \nabla.$$

This is a zeroth order differential operator and can be regarded as an endomorphism of $\Lambda^p T^*M$. The matrix of this endomorphism is linear in the second derivatives of the metric.

The formulas derived in [4] express $a_n(D_p^4)$ in terms of the covariant derivatives of the tensors R_{ijkl} , W_{ij}^p , $\mathcal{E}_p$ for $n = 0, 1, 2, 3$:

Theorem 3.2 : Let $\binom{4}{p}$ be the binomial coefficient.

$$(a) \quad a_0(D_p^4) = (4\pi)^{-2} \binom{4}{p},$$

$$(b) \quad a_1(D_p^4) = \binom{4}{p} a_1(D_0^4) + (4\pi)^{-2} \text{Tr}(\mathcal{E}_p),$$

$$(c) \quad a_2(D_p^4) = \binom{4}{p} a_2(D_0^4) + (4\pi)^{-2} \text{Tr} \left\{ \frac{1}{2} \mathcal{E}_p \mathcal{E}_p + \frac{1}{6} \mathcal{E}_{p;kk} \right. \\ \left. + \frac{1}{12} W_{ij}^p W_{ij}^p - \frac{1}{6} R_{ijij} \mathcal{E}_p \right\}$$

$$(d) \quad a_3(D_p^4) = \binom{4}{p} a_3(D_0^4) + (4\pi)^{-2} \text{Tr} \left\{ \frac{1}{60} \mathcal{E}_{p;jj} + \frac{1}{12} \mathcal{E}_{p;k} \mathcal{E}_{p;k} + \frac{1}{6} \mathcal{E}_{p;kk} \mathcal{E}_p \right. \\ \left. + \frac{1}{30} W_{ij;kk}^p W_{ij}^p + \frac{1}{45} W_{ij;k}^p W_{ij;k}^p + \frac{1}{180} W_{ij;i}^p W_{kk;k}^p \right. \\ \left. - \frac{1}{36} R_{ijij} \mathcal{E}_{p;kk} - \frac{1}{90} R_{ijik} \mathcal{E}_{p;jk} - \frac{1}{30} R_{ijij;k} \mathcal{E}_{p;k} \right. \\ \left. - \frac{1}{30} R_{ijij;kk} \mathcal{E}_p \right\} + \text{terms in the 2-jets of the metric.}$$

Because the alternating sum of the binomial coefficients vanishes and because $a_1^4 = 0$ by theorem 1.2 (a),

$$0 = a_1^4 = \sum_p (-1)^p a_1(D_p^4) = (4\pi)^{-2} \sum_p (-1)^p \text{Tr}(\mathcal{E}_p).$$

This shows that :

$$\sum_p (-1)^p \text{Tr}(\mathcal{E}_{p;kk}) = \{ \sum_p (-1)^p \text{Tr}(\mathcal{E}_p) \}_{;kk} = 0,$$

$$\sum_p (-1)^p \text{Tr}(R_{ijij} \mathcal{E}_p) = R_{ijij} \sum_p (-1)^p \text{Tr}(\mathcal{E}_p) = 0,$$

Using the multiplicative properties of the de Rham complex, one can show that both

$$\sum_p (-1)^p \text{Tr}(\mathcal{E}_p \mathcal{E}_p) \quad \text{and} \quad \sum_p (-1)^p \text{Tr}(W_{ij}^p W_{ij}^p),$$

belong to $\mathcal{K}_4^4$. If $g_{ab|cd}$ appears in the matrix W_{ij}^p , then both the indices i and j appear in $g_{ab|cd}$. This shows that the monomial $g_{11/22} g_{33/44}$ does not appear in $\sum_p (-1)^p \text{Tr}(W_{ij}^p W_{ij}^p)$. Since the linear functional $c(g_{11/22} g_{33/44}, P)$ separates $\mathcal{K}_4^4$, theorem 2.5 implies $\sum_p (-1)^p \text{Tr}(W_{ij}^p W_{ij}^p) = 0$. By theorem 1.2 (b), $E_2 = a_2^4$. Therefore

$$E_2 = \sum_p (-1)^p a_2(D_p^4) = (4\pi)^{-2} \sum_p \text{Tr}(\frac{1}{2} \mathcal{E}_p \mathcal{E}_p).$$

To compute c_1 and c_2 we ignore terms in the 2-jets of the metric. Because

$$\begin{aligned} \sum_p (-1)^p \text{Tr}(\mathcal{E}_p)_{;ijj} &= \{ \sum_p (-1)^p \text{Tr}(\mathcal{E}_p) \}_{;ijj} = 0, \\ \sum_p (-1)^p \text{Tr}(R_{ijk} \mathcal{E}_p)_{;jh} &= R_{ijk} \{ \sum_p (-1)^p \text{Tr}(\mathcal{E}_p) \}_{;jh} = 0, \\ \sum_p (-1)^p \text{Tr}(R_{ijj} \mathcal{E}_p)_{;kk} &= R_{ijj} \{ \sum_p (-1)^p \text{Tr}(\mathcal{E}_p) \}_{;kk} = 0, \\ \sum_p (-1)^p \text{Tr}(R_{ijj;k} \mathcal{E}_p)_{;k} &= R_{ijj;k} \{ \sum_p (-1)^p \text{Tr}(\mathcal{E}_p) \}_{;k} = 0, \\ \sum_p (-1)^p \text{Tr}(R_{ijj;kk} \mathcal{E}_p) &= R_{ijj;kk} \sum_p (-1)^p \text{Tr}(\mathcal{E}_p) = 0, \end{aligned}$$

the formula for a_3^4 reduces to :

$$\begin{aligned} a_3^4 &= (4\pi)^{-2} \sum_p (-1)^p \text{Tr} \left(\frac{1}{12} \mathcal{E}_p{}_{;k} \mathcal{E}_p{}_{;k} + \frac{1}{6} \mathcal{E}_p{}_{;kk} \mathcal{E}_p + \frac{1}{45} W_{ij;k}^p W_{ij;k}^p \right. \\ &\quad \left. + \frac{1}{180} W_{ij;i}^p W_{ik;k}^p + \frac{1}{30} W_{ij;kk}^p W_{ij}^p \right) + \text{terms in the 2-jets of the metric.} \end{aligned}$$

If $g_{ab|cd}$ appears in the matrix which represents $W_{ij;k}^p$, then both the indices i and j appear in $g_{ab|cd}$. This implies the monomial $g_{11/222} g_{33/244}$ does not appear in $\sum_p (-1)^p \text{Tr}(W_{ij;k}^p W_{ij;k}^p)$. Similarly, the monomial $g_{11/222} g_{33/44}$ does not appear in $\sum_p (-1)^p \text{Tr}(W_{ij;kk}^p W_{ij}^p)$.

If the variable $g_{11/222}$ appears in the matrix $W_{ij;j}^p$, then $j=2$. Since W_{ij}^p is skew-symmetric $i \neq j$. This implies $i=1$ and shows that if $g_{11/222} g_{ab|cd}$ is a monomial of $\text{Tr}(W_{ij;j}^p W_{ik;k}^p)$, then the index '1' appears in $g_{ab|cd}$. Therefore $g_{11/222} g_{33/244}$ is not a monomial of $\sum_p (-1)^p \text{Tr}(W_{ij;j}^p W_{ik;k}^p)$. This shows:

$$a_3^4 = (4\pi)^{-2} \sum_p (-1)^p \text{Tr} \left(\frac{1}{12} \mathcal{E}_p{}_{;k} \mathcal{E}_p{}_{;k} + \frac{1}{6} \mathcal{E}_p{}_{;kk} \mathcal{E}_p \right) + \dots$$

where we have omitted any terms involving only the 2-jets of the metric.

We have chosen the normalising constants so $E_m = \Theta_m^1 + \Theta_m^2$. Therefore:

$$\begin{aligned} E_{2;k} &= \Theta_4^1 + \Theta_4^2 = (4\pi)^{-2} \sum_p (-1)^p \text{Tr}(\frac{1}{2} \mathcal{E}_p \mathcal{E}_p)_{;kk} \\ &= (4\pi)^{-2} \sum_p (-1)^p \text{Tr}(\mathcal{E}_p{}_{;k} \mathcal{E}_p{}_{;k} + \mathcal{E}_p{}_{;kk} \mathcal{E}_p). \end{aligned}$$

Therefore :

$$a_3^4 - \frac{1}{6} E_{2;kk} = a_3^4 - \frac{1}{6} \Theta_4^1 - \frac{1}{6} \Theta_4^2 = (4\pi)^{-2} \sum_p (-1)^p \text{Tr} \left(-\frac{1}{12} \mathcal{E}_{p;k} \mathcal{E}_{p;k} \right) + \dots$$

$$a_3^4 - \frac{1}{12} E_{2;kk} = a_3^4 - \frac{1}{12} \Theta_4^1 - \frac{1}{12} \Theta_4^2 = (4\pi)^{-2} \sum_p (-1)^p \text{Tr} \left(\frac{1}{12} \mathcal{E}_{p;kk} \mathcal{E}_p \right) + \dots$$

Because $a_3^4 - 1/6 \Theta_4^1$ contains no terms in the 4-jets of the metric, $c_1 = 1/6$. Because $a_3^4 - 1/12 \Theta_4^2$ contains no terms which are quadratic in the 3-jets of the metric $c_2 = 1/12$. We may therefore express

$$a_n^m = \frac{1}{6} \Theta_m^1 + \frac{1}{12} \Theta_m^2 + \sum_{k>2} c_k \Theta_m^k.$$

Let $\tau(m) = 1/\{(-8\pi)^{\frac{1}{2}m} (\frac{1}{2}m - 1)!\}$. We have defined

$$\Phi_m = \tau(m) \sum_{I,J,k} \varepsilon(I, J) R_{i_1 i_2 j_1 k; k j_2} R_{i_2 i_4 j_3 j_4} \dots R_{i_{m-1} i_m j_{m-1} j_m},$$

$$\Theta_m^1 = \tau(m) \sum_{I,J,k} \varepsilon(I, J) R_{i_1 i_2 i_3 j_2; k k} R_{i_3 i_4 j_3 j_4} \dots R_{i_{m-1} i_m j_{m-1} j_m},$$

$$\Theta_m^3 = \tau(m) \sum_{I,J,k,l} \varepsilon(I, J) R_{k i_1 k j_1} R_{i_2 i_3 j_2} R_{i_3 i_4 j_3 j_4} \dots R_{i_{m-1} i_m j_{m-1} j_m},$$

$$\Theta_m^5 = \tau(m) \sum_{I,J,k,l} \varepsilon(I, J) R_{k i_1 i_2} R_{k i_3 j_2} R_{i_2 i_4 j_3 j_4} \dots R_{i_{m-1} i_m j_{m-1} j_m}.$$

By using the various curvature identities, we compute that:

$$\Phi_m = \frac{1}{2} \Theta_m^1 + b_3 \Theta_m^3 + b_5 \Theta_m^5,$$

where b_1 and b_2 are universal constants independent of m . Therefore:

$$\begin{aligned} a_n^m &= \frac{1}{6} \Theta_m^1 + \frac{1}{12} \Theta_m^2 + \sum_{k>2} c_k \Theta_m^k = \frac{1}{12} \Theta_m^1 + \frac{1}{12} (\Theta_m^1 + \Theta_m^2) + \sum_{k>2} c_k \Theta_m^k \\ &= \frac{1}{6} \Phi_m + \frac{1}{12} E_{m;kk} + \sum_{k>2} c'_k \Theta_m^k. \end{aligned}$$

It is clear $E_{m;kk}$ is a divergence. We show Φ_m is a divergence by computing:

$$\begin{aligned} &\{ \tau(m) \sum_{I,J,k} \varepsilon(I, J) R_{i_1 i_2 j_1 k; k j_2} R_{i_2 i_4 j_3 j_4} \dots R_{i_{m-1} i_m j_{m-1} j_m} \}_{j_2} \\ &= \Phi_m + (\frac{1}{2}m - 1) \tau(m) \sum_{I,J,k} \varepsilon(I, J) R_{i_1 i_2 j_2 k; k} R_{i_2 i_4 j_3 j_4; j_2} \dots R_{i_{m-1} i_m j_{m-1} j_m}. \end{aligned}$$

By the second Bianci identity $R_{i_2 i_4 j_3 j_4; j_2} = 0$ when we alternate over $(j_3 j_4 j_2)$. This shows Φ_m is a divergence so

$$\int_M \Phi_m(x, G) |d\text{vol}| = \int_M E_{m;kk}(x, G) |d\text{vol}| = \int_M a_n^m(x, G) |d\text{vol}| = 0.$$

This establishes that $\int_M \sum_{k>2} c'_k \Theta_m^k |d\text{vol}| = 0$.

The following lemma proves $c'_k = 0$ for $k > 2$ and completes the proof of theorem 1.3:

Lemma 3.3

Let $P \in \mathcal{K}_6^8$ and suppose that (a) $\int_M P(G) |dvol| = 0$ for every six-dimensional Riemannian manifold M and that (b) $P = \sum_{k \geq 2} a_k \Theta_k^4$. Then $P = 0$.

Proof: We will first show that $a_3 = 0$.

Let $Q = a_3 \Theta_4^3 + a_4 \Theta_4^4 + a_5 \Theta_4^5 \in \mathcal{K}_6^4$

and let G_0 be the standard metric on S^2 . If c is a positive constant, $c^2 G_0$ is the metric on the sphere of radius c in R^3 . S^2 with the metric $c^2 G_0$ is a homogeneous space so $E_2(c^2 G_0)(x_2) = c^{-2} E_2(G_0)$ and $\Theta_2^4(c^2 G_0)(x_2) = c^{-4} \Theta_2^4(G_0)$ is independent of the point $x_2 \in S^2$. Let $M^6 = M_1^4 \times S^2$ with the metric $G = G_1 \times (c^2 G_0)$.

Then $P(G)(x) = Q(G_1)(x_1) \cdot c^{-2} E_2(G_0) + E_4(G_1)(x_1) \cdot c^{-4} \sum_i a_i \Theta_2^i(G_0)$.

Since $|dvol|(G) = |dvol|(G_1) \cdot c^2 |dvol(G_0)|$,

we integrate over M^6 to show that

$$0 = \int_{M_1^4} Q(G_1) |dvol|(G_1) \cdot \chi(S^2) + c^{-2} \chi(M_1^4) \text{vol}(S^2(G_0)) \sum_i a_i \Theta_2^i(G_0).$$

We take the limit as $c \rightarrow \infty$ to show $\int_{M_1^4} Q(G_1) |dvol|(G_1) = 0$ as well.

Let $M_1^4 = S^1 \times S^1 \times S^1 \times S^1$ to be the four-dimensional torus. Let $x = (x_1, x_2, x_3, x_4)$ be the usual periodic parameters, let $\varepsilon > 0$ and let $h(x_2)$ be a function of the second parameter alone. We let $G(\varepsilon)$ be the conformally flat metric with

$$ds^2 = \exp(-\varepsilon h) (dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2).$$

We compute that:

$$R_{i2i2} = R_{2i2i} = -R_{2i4i} = -R_{i22i} = -\varepsilon/2 h_{i22} + O(\varepsilon^2)$$

$$R_{ijkl} = O(\varepsilon^2) \text{ otherwise.}$$

Consequently:

$$\Theta_4^3 = \lambda \varepsilon^3 \exp(3\varepsilon h) (h_{i22})^3 + O(\varepsilon^4) \text{ for some constant } \lambda \neq 0$$

$$\Theta_4^4 = O(\varepsilon^4) \text{ and } \Theta_4^5 = O(\varepsilon^4).$$

The factor of $\exp(3\varepsilon h)$ is introduced since these coordinates are not normalised. Suppose that $a_3 \neq 0$.

$$|dvol|(G(\varepsilon)) = \exp(-2\varepsilon h) dx_1 dx_2 dx_3 dx_4.$$

Since h depends only on x_2 , $Q(G(\varepsilon))(x)$ only depends on x_2 and

$$0 = \int_M Q(G(\varepsilon)) |dvol| = \lambda a_3 \varepsilon^3 \int (h_{i22})^3 \exp(\varepsilon h) dx_2 + O(\varepsilon^4).$$

By multiplying by $(\varepsilon^3 \lambda a_3)^{-1}$ and taking the limit as $\varepsilon \rightarrow 0$, we show that

$$0 = \int (h'')^3 dx_2$$

for every periodic function of one variable h . Since this is not true, $a_3 = 0$.

Let G_0 be the standard metric on S^3 and let $M^6 = S^3 \times S^3$ with the product metric $G = G_0 \times G_0$. Then :

$$\Theta_6^4(G) = 0, \quad \Theta_6^5(G) = 0,$$

$$\Theta_6^6(G)(x) = \Theta_6^6(G) = \Psi_3(G_0)^2 \text{ is independent of } x.$$

Therefore

$$\int_M P(G) |d\text{vol}| = \Psi_3(G_0)^2 a_6 \text{vol}(M^6) = 0.$$

To show $a_6 = 0$, it suffices to show that $\Psi_3(G_0) \neq 0$. On any sphere S^k with the standard metric, the only non-zero curvatures are

$$R_{ijij} = -R_{ijji} = -1.$$

We defined

$$\Psi_3 = c \sum_{I,J,k} \varepsilon(I, J) R_{k i_1 j_2 i_3} R_{k j_1 i_2 i_3}$$

where $c \neq 0$ was a normalising constant. Therefore $\Psi_3 = -24c \neq 0$ on S^3 so $a_6 = 0$.

This implies that $P = a_4 \Theta_6^4 + a_5 \Theta_6^5$ so $Q = a_4 \Theta_4^4 + a_5 \Theta_4^5$. The manifolds $S^2 \times S^2$ and S^4 with the standard metrics are homogeneous. Since $\int Q(G) |d\text{vol}| = 0$, $Q(G) = 0$ pointwise for these manifolds. We defined

$$\Theta_4^4 = c \sum_{I,J,k,l} \varepsilon(I, J) R_{k l k l} R_{i_1 i_2 j_1 j_2} R_{i_3 i_4 j_3 j_4},$$

$$\Theta_4^5 = d \sum_{I,J,k,l} \varepsilon(I, J) R_{k l i_1 i_2} R_{k l j_1 j_2} R_{i_3 i_4 j_3 j_4},$$

where c and d are non-zero normalising constants. We compute that :

$$\Theta_4^4(S^4) = -48(4!)c,$$

$$\Theta_4^5(S^4) = -8(4!)d,$$

$$\Theta_4^4(S^2 \times S^2) = -128c,$$

$$\Theta_4^5(S^2 \times S^2) = -64d.$$

This yields the system of equations

$$48(ca_4) + 8(da_5) = 0,$$

$$128(ca_4) + 64(da_5) = 0.$$

Since c and d are non-zero and since this system of equations is non-singular, $a_4 = a_5 = 0$. This completes the proof of the lemma and thereby of theorem 1.3.

McKean and Singer [8] showed that $a_n^2 = (P_n^2)_{;kk}$ for $n = 1$ where P_n^2 is a scalar invariant. We showed [6] that in general $a_n^m = \text{div}(Q_n^m)$ for $2n = m$ where Q_n^m was 1-form valued. It is natural to ask if we can represent $a_n^m(P_n^m)_{;kk}$ in general.

The answer to this question is no for $m > 2$ by theorem 1.3. We show this as follows :

Suppose $a_n^m = (P_n^m)_{;kk}$ for $2n = m + 2$. Since $r(a_n^m) = 0$, $r(P_n^m)_{;kk} = 0$. This implies that $r(P_n^m) = 0$. P_n^m is a scalar invariant of order $2n - 2 = m$ so therefore $P_n^m = cE_m$. This would imply that $a_n^m = cE_{m;kk} = c\Theta_m^1 + c\Theta_m^2$. Since the coefficient of Θ_m^1 in a_n^m is $1/6$ and the coefficient of Θ_m^2 in a_n^m is $1/12$, this fails. This shows that the divergence result of [6] is the strongest possible result in general and that such a formal Hodge decomposition theorem fails.

In the course of proving theorem 2.5, we showed $\dim(\mathfrak{K}_{m+2}^m) = 6$ for $m \geq 6$ and m even, that $\dim(\mathfrak{K}_{m+2}^{m+1}) = 1$ for $m \geq 2$ and m even, and that $\dim(\mathfrak{K}_m^m) = 1$ for $m \geq 2$ and m even. There is a general stability theorem concerning $\dim(\mathfrak{K}_{m+k}^m)$ which is true for all integers k and m sufficiently large which we discuss for the sake of completeness. Let M_1 be a Riemannian manifold of dimension $m - 2$ and let M_2 be a Riemannian manifold of dimension 2. Let $c \neq 0$ and let $G = G_1 \times c^2 G_2$ be the product metric. Let $P \in \mathfrak{K}_{m+k}^m$ and decompose

$$P(G)(x) = \sum_i c^{-2i} P'_i(G_1)(x_1) P''_i(G_2)(x_2),$$

where P'_i is homogeneous of order $2i$ in the derivatives of the metric G_2 and where P''_i is homogeneous of order $m + k - 2i$ in the derivatives of the metric G_1 . It is easily verified that $P'_i \in \mathfrak{K}_{m+k-2i}^{m-2}$ and that $P''_i \in \mathfrak{K}_{2i}^2$. Since $\mathcal{P}_0^2$ just consists of the constants, $\mathfrak{K}_0^2 = 0$ so $P'_0 = P''_0 = 0$. $\mathfrak{K}_2^2$ is one-dimensional and is spanned by E_2 .

This permits us to decompose

$$P(G)(x) = c^{-2} P'_1(G_1)(x_1) E_2(G_2)(x_2) + O(c^{-4}).$$

We define the linear map

$$\tau : \mathfrak{K}_{m+k}^m \rightarrow \mathfrak{K}_{m+k-2}^{m-2} \text{ by } \tau(P) = P'_1.$$

For example, in the proof of lemma 3.3, we defined $Q(P) \in \mathfrak{K}_6^4$ given $P \in \mathfrak{K}_8^6$. In this notation, $Q(P) = \tau(P)$. Furthermore,

$$\tau(a_n^m) = a_{n-1}^{m-2}$$

$$\tau(E_m) = E_{m-2}$$

$$\tau(\Psi_m) = \Psi_{m-2}$$

$$\tau(\Theta_m^i) = \Theta_{m-2}^i.$$

We conjecture that the following is true :

Conjecture 3.4

(a) τ is always surjective.

(b) τ is an isomorphism provided that $m \geq m(k)$ where $m(0) = 4$ and $m(k) = 3k + 2$ for $k > 0$. [We define $m(0) = 4$ to avoid having to consider the spaces $\mathfrak{K}_0^0$.]

Remark : Theorem 2.5 together with the above computations proves this conjecture if $k = 0, 1, 2$. It is fairly easy to adapt the methods used in the proof of theorem 2.5 to show that τ is injective if $m \geq 8k + 2$. If we let $m = 3k$ and define

$$A = \sum_{I, J, p_1, \dots, p_k \in (I, J)} R_{p_1, i_1, j_2, j_3} R_{p_1, i_1, i_2, i_3} R_{p_2, i_4, j_5, j_6} R_{p_2, i_4, i_5, i_6} \dots$$

then $A \neq 0 \in \mathcal{K}_{4k}^k$ and $\tau(A) = 0$.

Irrespective of this conjecture, we can prove that τ is injective for $m \geq 8k + 2$ so $\lim_{m \rightarrow \infty} \dim(\mathcal{K}_{m+k}^m) = \mathcal{E}'(k)$ is well-defined (where we take the limit over $m+k$ even). Theorem 2.5 implies that $\mathcal{E}(0) = 1$, $\mathcal{E}(1) = 1$, and $\mathcal{E}(2) = 6$. It would be interesting to obtain further values of $\mathcal{E}(k)$ and to investigate further the structure of the spaces $\mathcal{K}_{m+k}^m$. This might make possible the computation of still higher terms in the asymptotic expansion of the heat equation for the de Rham complex.

References

- [1] Atiyah M, Bott R and Patodi V K 1973 On the heat equation and the index theorem; *Invent. Math.* **19** 279–330
- [2] Chern S 1944 A simple intrinsic proof of the Gauss-Bonnet formula for closed Riemannian manifolds; *Ann. Math.* **45** 747–752
- [3] Gilkey P 1972 *Curvature and the eigenvalues of the Laplacian for geometrical elliptic complex*, Thesis, Harvard University
- [4] Gilkey P 1973 Curvature and the eigenvalues of the Laplacian for elliptic complexes; *Adv. Math.* **10** 344–382
- [5] Gilkey P 1975a The spectral geometry of a Riemannian manifold; *J. Diff. Geo.* **10** 601–618
- [6] Gilkey P 1975b Local invariants of a pseudo-Riemannian manifold; *Math. Scand.* **36** 109–130
- [7] Gilkey P 1979 Recursion relations and the asymptotic behaviour of the eigenvalues of the Laplacian; *Compositio Mathematica* **38** 201
- [8] McKean H P and Singer I M 1967 Curvature and the eigenvalues of the Laplacian; *J. Diff. Geo.* **1** 43–69
- [9] Patodi V K 1970 Curvature and the fundamental solution of the heat equation; *J. Indian Math. Soc.* **34** 269–285
- [10] Patodi V K 1971 Curvature and the eigenforms of the Laplace operator; *J. Diff. Geo.* **5** 233–249
- [11] Sakai T 1971 On the eigenvalues of the Laplacian and curvature of Riemannian manifold *Tohoku Math. J.* **23** 589–603
- [12] Weyl H 1946 *The classical groups* (Princeton : University Press)

Fine Hall-Box 37,
Princeton,
New Jersey 08540, USA

Units of Hill curves

By

H P McKEAN*

Dedication

This paper is dedicated to the memory of V. K. Patodi. One of Patodi's interests was the heat equation on manifolds and what would now be called *trace formulas* relating the spectrum of the Laplace operator to the geometry. Trace formulas play an important role in the recent investigations of one-dimensional iso-spectral problems in their relation to nonlinear waves and lattices. This development would undoubtedly have fascinated him. The present paper is a small chapter therein, its purpose being to relate the periodicity of isospectral flows to the periods of differentials of the second kind on certain algebraic curves of infinite genus, and to the existence of appropriate units in the corresponding function fields. I hope that it would have pleased him.

1. Introduction

This paper is a continuation of McKean-Trubowitz [5], [6], dealing with further aspects of the function theory of a real hyperelliptic curve $\mathbf{K}$ of genus $g \leq \infty$, having a unit e in its function field of the form $e(\mathbf{p}) \sim \exp \sqrt{-\lambda}$ near ∞ . This type of curve arises from the study of Hill's operator $Q = -d^2/dx^2 + q(x)$ with real infinitely differentiable potential of period 1. The existence of the unit expresses the periodicity of Q ; its importance was pointed out by Marcenko Ostrovskii [3]; see also Krichever [2]. The branch points of the curve comprise the simple periodic spectrum of Q , and the class of all Q' having the same spectrum as Q may be identified as the real part $\mathbf{J}$ of the Jacobi variety of the curve; moreover, the isospectral motions of Q , such as translation $\partial q/\partial t = \partial q/\partial x$, Korteweg-de Vries

$$\partial q/\partial t = 3q \partial q/\partial x - (1/2) \partial^3 q/\partial^3 x,$$

etc., are reflected in straight-line motions at constant speed on the Jacobian. The subject of this paper is the periodicity of the higher flows and its connection with

* The research for this paper was done at the Courant Institute of Mathematical Sciences. The author gratefully acknowledges the support of the National Science Foundation under Grant No. NSF-MCS 76-97039. Reproduction in whole or in part is permitted for any purpose of the U.S. Govt.

higher units of the function field of $\mathbf{K}$. The plan of the paper is as follows. § 2 is preliminary: it explains the connection between Q and the Jacobi variety as in McKean-Trubowitz [5]. § 3 deals with the spatial periodicity of Q and its expression in terms of the periods of $\mathbf{J}$, extending to $g = \infty$, a theorem of McKean-van Moerbeke [4]. § 4 deals with the higher flows, expressing the corresponding directions on $\mathbf{J}$ in terms of the periods of differentials of the second kind on $\mathbf{K}$. § 5 presents the connection with the higher units of the function field. § 6 has been kindly contributed by van Moerbeke [7]: it deals with the higher units and flows for the simpler case of curves of finite genus arising from the periodic Toda lattice and may profitably be read first to provide an orientation.

2. Hill's equation. The Jacobi variety

The following material may be found in McKean-Trubowitz [5]. Let Q be as above and let

$$\lambda_0 < \lambda_1 \leq \lambda_2 < \lambda_3 \leq \lambda_4 < \dots \uparrow \infty$$

be the *periodic spectrum* of Q determined by the periodic and anti-periodic solutions of

$$Qf = \lambda f \quad [f(x+1) = \pm f(x)].$$

The eigenspace of λ_{2i-1} or λ_{2i} is one-dimensional if $\lambda_{2i-1} < \lambda_{2i}$ and two-dimensional if $\lambda_{2i-1} = \lambda_{2i}$. The associated eigenfunctions f_{2i-1} and f_{2i} are alternately anti-periodic and periodic by pairs, f_0 being periodic and root-free, while f_{2i-1} and f_{2i} have each i roots per period $0 \leq x < 1$. λ_{2j-1} and λ_{2j} have a common development

$$j^2 \pi^2 + k_0 + k_4 j^{-4} + k_2 j^{-2} + \dots$$

The coefficient k_0 is the integral of $q(x)$ over a period; it is taken to be 0 with a view to the simplicity of certain formulae. The spectrum may be described as the roots of $\Delta(\lambda) = \pm 1$, the *discriminant* Δ being half the trace $y_1(1, \lambda) + y_2'(1, \lambda)$ of the monodromy matrix of Q formed from the solutions $y_1(x, \lambda)$ [$y_2(x, \lambda)$] of $Qy = \lambda y$ with

$$y_1(0) = 1, y_1'(0) = 0 \quad [y_2(0) = 0, y_2'(0) = 1].$$

The class M of Q with *fixed* periodic spectrum is a compact torus of $g \leq \infty$ dimensions, $2g+1$ being the number of simple eigenvalues of Q . This figure may be visualised as the real part $\mathbf{J}$ of the Jacobi variety of the hyperelliptic irrationality

$$y(\lambda) = [(\lambda_0 - \lambda) \prod_{j=1}^{\infty} (j\pi)^{-4} (\lambda_{2j-1} - \lambda) (\lambda_{2j} - \lambda)]^{1/2} = [\Delta^2(\lambda) - 1]^{1/2}$$

properly desingularised, i.e., with double roots of the radicand removed. This will now be explained in more detail. Let the numbers $\mu_1 < \mu_2 < \dots \uparrow \infty$ be

the auxiliary spectrum of Q obtained by solving $Qf = \mu f$ with $f(0) = f(1) = 0$. The interval $[\lambda_{2i-1}, \lambda_{2i}]$ contains precisely one such μ_i , and the corresponding eigenfunction is proportional to $y_2(x, \mu_i)$. Borg [1] proved that Q is uniquely specified by the auxiliary spectrum together with the norming constants (the superscript spot signifies $\partial/\partial\mu$)

$$\int_0^1 y_2^2(x, \mu_i) dx = y_2^{\bullet}(1, \mu_i) \times \frac{1}{2} \{ \Delta(\mu_i) - [\Delta^2(\mu_i) - 1]^{1/2} \}.$$

Now for fixed periodic spectrum, the knowledge of the i th norming constant adds only the information of the signature of the radical $[\Delta^2(\mu_i) - 1]^{1/2} = y(\mu_i)$. The latter is immaterial if $\mu_i = \lambda_{2i-1}$ or λ_{2i} as $y(\mu)$ vanishes there. This suggests replacing the interval $[\lambda_{2i-1}, \lambda_{2i}]$ by two banks distinguished by that signature, forming a circle C_i of points $\mathbf{p}_i = [\mu_i, y(\mu_i)]$.

Borg's theorem now states that the map

$$Q \rightarrow (\mathbf{p}_1, \mathbf{p}_2, \dots) \in C_1 \times C_2 \times \dots$$

is $1:1$; it is, in fact, onto, confirming that M is a g -dimensional torus.

Now the radical $y(\lambda) = [\Delta^2(\lambda) - 1]^{1/2}$ defines a singular curve forming 2-sheeted covering of the sphere, punctured at ∞ if $g = \infty$: each pair of *simple* eigenvalues $\lambda_{2i-1} < \lambda_{2i}$ contributes a handle, while the *double* eigenvalues $\lambda_{2i-1} = \lambda_{2i}$ contribute unimportant double points. $\mathbf{K}$ denotes the desingularised curve; it is of genus g , $2g + 1$ being the cardinality of the *simple* periodic spectrum of Q . Consider the class of integral functions $\phi(\lambda)$ of order $1/2$ and type ≤ 1 such that

$$(a) \quad \phi(\lambda^*) = \phi^*(\lambda),$$

$$(b) \quad \int_0^\infty |\phi(\lambda)|^2 \lambda^{3/2} d\lambda < \infty,$$

$$(c) \quad \phi(\lambda) \text{ vanishes at double eigenvalues } \lambda_{2i-1} = \lambda_{2i}, \text{ and}$$

$$(d) \text{ the values of } \phi(\lambda) \text{ in the intervals } [\lambda_{2i-1}, \lambda_{2i}] \text{ vanish rapidly as } i \uparrow \infty.$$

The forms $\omega = \phi(\lambda) d\lambda/y(\lambda)$ are declared to be the differentials of the first kind [DFK] of the desingularised curve $\mathbf{K}$; in fact, if $g < \infty$, they reduce to the classical ones. To define the real part $\mathbf{J}$ of the Jacobi variety of $\mathbf{K}$, let the divisor $\mathbf{d} : \mathbf{p}_i (i = 1, 2, \dots)$ be in *real position* on $\mathbf{K}$, meaning that the projection of $\mathbf{p}_i$ lies in $[\lambda_{2i-1}, \lambda_{2i}]$, let $\mathbf{o}_i = [\lambda_{2i-1}, 0]$, and form the sum

$$\mathbf{x}(\omega) = \sum_{i=1}^{\infty} \int_{\mathbf{o}_i}^{\mathbf{p}_i} \omega \equiv \int \omega,$$

for differentials of the first kind. The path of integration $\mathbf{o}_i \mathbf{p}_i$ is taken in real position, and it is required that the number n_i of full revolutions it makes about C_i does not grow faster than a power of i . This recipe provides a map from the divisor $\mathbf{d}$ onto the dual space (DFK)' of DFK, the latter being provided with a self-evident topology. (DFK)' is now factored by the lattice $\mathbf{L}$ of periods $\mathbf{x}$ arising from closed paths of integration $[\mathbf{p}_i = \mathbf{o}_i]$, and a faithful copy of $C_1 \times C_2 \times \dots$, alias M , is obtained. The factor space $\mathbf{J} = (\text{DFK})'/\mathbf{L}$ is the real part of the Jacobian cited before.

The fact that Q is of period 1 has a peculiar expression in terms of the Jacobi variety. Let Q be translated by an amount $0 \leq x < 1$. This does not change the periodic spectrum, i.e., translation preserves M , but it *does* change the auxiliary spectrum μ_i ($i = 1, 2, \dots$), so that these numbers now depend upon x ; in fact (the superscript dot denotes $\partial/\partial\lambda$)

$$d\mu_i/dx = 2y(\mu_i)/D^*(\mu_i) \quad (i = 1, 2, \dots)$$

$$D(\lambda) = \prod_{j=1}^{\infty} (j\pi)^{-2} (\mu_j - \lambda),$$

so that the direction $\mathbf{x} \in (\text{DFK})'$ corresponding, at $\mathbf{x} \in \mathbf{J}$, to the infinitesimal translation is

$$\mathbf{x}_1(\omega) = \frac{d}{dx} \int_0^p \omega = \sum_{i=1}^{\infty} \frac{\phi(\mu_i)}{y(\mu_i)} \frac{d\mu_i}{dx} = 2 \sum_{i=1}^{\infty} \frac{\phi(\mu_i)}{D^*(\mu_i)}.$$

The expression can be simplified: The function $\phi(\lambda)$ figuring in ω can be interpolated off any series of points

$$\mu_i \in [\lambda_{2i-1}, \lambda_{2i}] \quad (i = 1, 2, \dots)$$

by the recipe

$$\phi(\lambda) = \sum_{i=1}^{\infty} \phi(\mu_i) \prod_{j \neq i} \frac{\lambda - \mu_j}{\mu_i - \mu_j} = \sum_{i=1}^{\infty} \frac{\phi(\mu_i)}{\lambda - \mu_i} \frac{D(\lambda)}{D^*(\mu_i)},$$

$$\text{so } \frac{1}{2} \mathbf{x}_1(\omega) = \mathbf{c}_0(\omega) = \lim_{\lambda \downarrow -\infty} \frac{\lambda \phi(\lambda)}{D(\lambda)} = \lim_{\lambda \downarrow -\infty} - \frac{(-\lambda)^{3/2} \phi(\lambda)}{\sin \sqrt{-\lambda}}.$$

This proves two things: (1) $\mathbf{x}_1$ does not depend upon the place $\mathbf{x} \in \mathbf{J}$, so that, on $\mathbf{J}$, translation is a straight-line motion at constant speed, and (2)

$$\mathbf{x}_1(\omega) = \int_0^1 \sum_{i=1}^{\infty} \frac{\phi(\mu_i)}{y(\mu_i)} d\mu_i = \sum_{i=1}^{\infty} i \mathbf{a}_i(\omega),$$

$\mathbf{a}_i(\omega)$ being the i th (real) period of $\omega \in \text{DFK}$ obtained by anti-clockwise integration about $[\lambda_{2i-1}, \lambda_{2i}]$. The number i appearing under the sum is the number of full revolutions of $\mathbf{p}_i$ in one period $0 \leq x < 1$, alias the number of roots per period of f_{2i-1} or f_{2i} . The connection comes about by virtue of the fact that the eigenfunction for simple $\lambda = \lambda_{2i-1}$ or λ_{2i} is proportional to $\sqrt{D(\lambda)}$; note that $\mathbf{a}_i(\omega)$ drops out of the sum if $\lambda_{2i-1} = \lambda_{2i}$ is *double* thanks to the root of $\phi(\lambda)$ at that place. The formula expresses the fact that *the direction of translation*

$$\mathbf{x}_1 = 2 \mathbf{c}_0 \in (\text{DFK})' \text{ is a period } \sum_{i=1}^{\infty} i \mathbf{a}_i \in \mathbf{L}.$$

3. Hill's equation: Periodicity continued

Let $\lambda_0 < \lambda_1 \leq \lambda_2 < \lambda_3 \leq \lambda_4 < \dots \uparrow \infty$ to be fixed and let λ_{2j-1} and λ_{2j} have a common development

$$j^2 \pi^2 + k_1 j^{-2} + k_2 j^{-4} + \dots \quad (j \uparrow \infty).$$

The theory of the curve $\mathbf{K}$ defined by the irrationality $\mathbf{y}(\lambda)$ can be developed just as in the preceding article. The purpose of the present article is to prove that (see [4] for the case of $< \infty$) the periodicity condition $2\mathbf{c}_0 \in \mathbf{L}$ is not only necessary but also sufficient that $\lambda_0 < \lambda_1 \leq \lambda_2 < \dots$ be the periodic spectrum of some Hill's operator $Q = -d^2/dx^2 + q(x)$ of period 1; see [4] for the case $g < \infty$. The proof is indicated in a series of simple steps. For simplicity, suppose that $\lambda_{2i-1} < \lambda_{2i}$ for every i .

Step 1. The differential equation

$$d\mu_i/dx = 2\mathbf{y}(\mu_i)/D^*(\mu_i) \quad (i = 1, 2, \dots)$$

is pulled out of air and it is proved that it has precisely one solution

$$\mu_i(x) \in [\lambda_{2i-1}, \lambda_{2i}] \quad (i = 1, 2, \dots)$$

starting at $\mu_i(0) = \lambda_{2i-1}$, with the understanding that every time μ_i hits λ_{2i-1} or λ_{2i} , the signature of the radical $\mathbf{y}(\mu_i)$ flips and μ_i continues its motion without rest; also, $\mathbf{p}_i = [\mu_i, \mathbf{y}(\mu_i)]$ winds approximately i times about C_i for $i \uparrow \infty$. Trubowitz [10] may be consulted for details.

Step 2 is to prove that the above motion of μ_i ($i = 1, 2, \dots$) is of period 1.

Proof. $2\mathbf{c}_0 \in \mathbf{L}$ is used now. The motion is rectilinear at constant speed on $\mathbf{J}$, as in § 2, and the corresponding infinitesimal direction in $(\mathbf{DFK})'$ may be expressed either as

$$2\mathbf{c}_0(\omega) = \lim_{\lambda \downarrow -\infty} -(-\lambda)^{3/2} \phi(\lambda) / \sin \sqrt{-\lambda},$$

$$\text{or as } \mathbf{x}_1(\omega) = \sum_{i=1}^{\infty} \int_{\mathbf{p}_i(0)}^{\mathbf{p}_i(1)} \omega,$$

as you will. The identity of the two implies that $\mathbf{x}_1$ is a period of class $\mathbf{L}$. But the map from divisors $\mathbf{d}: \mathbf{p}_i$ ($i = 1, 2, \dots$) in real position on $\mathbf{K}$ to $\mathbf{J}$ is 1 : 1 modulo periods, so the paths of integration leading from $\mathbf{o}_i = \mathbf{p}_i(0)$ to $\mathbf{p}_i = \mathbf{p}_i(1)$, are closed and $\mathbf{p}_i(0) = \mathbf{p}_i(1)$, as required.

Step 3 is to observe that the function

$$q(x) = \lambda_0 + \sum_{i=1}^{\infty} [\lambda_{2i-1} + \lambda_{2i} - 2\mu_i(x)]$$

is infinitely differentiable and of period 1. The first point is proved as in Trubowitz [10]. The common development of λ_{2i-1} and λ_{2i} implies that

$$|\lambda_{2i-1} + \lambda_{2i} - 2\mu_i| \leq (\lambda_{2i} - \lambda_{2i-1})$$

vanishes rapidly as $i \uparrow \infty$, permitting indefinite differentiation under the sum. The second point is self-evident from Step 2.

Step 4 is to identify the periodic spectrum of $Q = -d^2 dx^2 + q(x)$ with

$$\lambda_0 < \lambda_1 \leq \lambda_2 < \dots$$

This is achieved much as in McKean-van Moerbeke [4]. Let

$$D(x) = \prod_{j=1}^{\infty} (j\pi)^{-2} [\mu_j(x) - \lambda].$$

A tiresome but elementary computation confirms that (the prime denotes $\partial/\partial x$)

$$A = \frac{1}{2} (lgD)'' + \frac{1}{4} [(lgD)']^2 - q + \lambda + y^2/D^2,$$

qua function of λ , is pole-free for fixed $0 \leq x < 1$. Besides, the first two pieces vanish at $\lambda = \infty$, while

$$\begin{aligned} y^2/D^2 &= (\lambda_0 - \lambda) \prod_{i=1}^{\infty} (\lambda_{2i} - \lambda) (\lambda_{2i-1} - \lambda) (\mu_i - \lambda)^{-2}, \\ &= (\lambda_0 - \lambda) \left[1 - \frac{1}{\lambda} \sum_{i=1}^{\infty} (\lambda_{2i-1} + \lambda_{2i} - 2\mu_i) + o\left(\frac{1}{\lambda^2}\right) \right], \\ &= \lambda_0 + \sum_{i=1}^{\infty} (\lambda_{2i-1} + \lambda_{2i} - 2\mu_i) - \lambda + o(1), \end{aligned}$$

near by, so that the whole is $o(1)$ by the formula of Step 3. This proves that $A = 0$, which is now used, as in McKean-van Moerbeke [4], to verify that the eigenfunctions $y_1(x, \lambda)$ and $y_2(x, \lambda)$ of Q can be expressed as

$$\begin{aligned} y_1(x, \lambda) &= [D(x)/D(0)]^{1/2} \times \left(\cos B - \frac{1}{2} D'(0) \frac{\sin B}{(-1)^{1/2} y} \right) \\ y_2(x, \lambda) &= [D(x)/D(0)]^{1/2} \frac{\sin B}{(-1)^{1/2} y}, \end{aligned}$$

$$\text{with } B = \sqrt{-1} y \int_0^x D^{-1},$$

$$\text{and } \lambda \notin [\lambda_{2i-1}, \lambda_{2i}] \quad (i = 1, 2, \dots).$$

The discriminant Δ of Q may now be expressed as $\cos B$, evaluated at $x = 1$, $y(\lambda)$ is imaginary in $\lambda_{2i-1} \leq \lambda \leq \lambda_{2i}$, so $-1 \leq \Delta(\lambda) \leq +1$ and no simple eigenvalues of Q are to be found there, the latter being the *simple* roots of $\Delta(\lambda) = \pm 1$. The proof is finished by use of the identity $A = 0$ to prove that if $\lambda_{2i-1} < \lambda_{2i}$, then for $\lambda = \lambda_{2i-1}$ or λ_{2i} , the function $f = \sqrt{D}$ is a periodic or anti-periodic eigenfunction of Q having the *same* number of roots in a period $0 \leq x < 1$ as the *common* number of visits of μ_i to λ_{2i-1} or λ_{2i} . This means that λ_{2i-1} and λ_{2i} are simple eigenvalues of Q , and that *no periodic spectrum lies between*.

4. Higher flows and differentials of the second kind

The tangent space to $\mathbf{J}$ can be identified with $(DFK)'$; it contains, but in general is not spanned by, the directions $\mathbf{x}_1$ corresponding to the equation of translation $\partial q/\partial t = \partial q/\partial x = X_1 q$, $\mathbf{x}_2$ corresponding to the equation of Korteweg-de Vries

$$\partial q/\partial t = 3q \partial q/\partial x - (1/2) \partial^3 q/\partial x^3 = X_2 q,$$

etc. (see [5]). The prescription for the higher flows is*

$$\partial q / \partial t = X_j q = D \partial H_j / \partial q$$

with $H_{j+1} q = (qD + Dq - (1/2) D^2) \partial H_j / \partial q$ ($j = 0, 1, 2, \dots$),

starting from $H_0 = \int_0^1 q(x) dx$;** as for X_1 , the direction $\mathbf{x}_j \in (\text{DFK})'$ corresponding to X_j is independent of the place $\mathbf{x} \in \mathbf{J}$ corresponding to Q , so that $\partial q / \partial t = X_j q$ induces on $\mathbf{J}$ a straight-line motion at constant speed.

McKean-Trubowitz [6] proved that if

$$\omega = \phi(\lambda) d\lambda / y(\lambda) \in \text{DFK},$$

and if $D(\lambda) = \prod_{j=1}^{\infty} (j\pi)^{-2} (\lambda_{2j-1} - \lambda)$,

then $\lambda \phi(\lambda) / D(\lambda)$ has a development

$$\sum_{j=0}^{\infty} c_j(\omega) \lambda^{-j},$$

on nice circles, i.e., large circles centred at λ_0 , avoiding the periodic spectrum by 1 or more; in this development

$$c_j(\omega) = \sum_{i=1}^{\infty} \frac{\phi(\lambda_{0i-1})}{D^*(\lambda_{2i-1})} \lambda_{2i-1}^j \quad (j = 0, 1, 2, \dots)$$

is an element of $(\text{DFK})'$. Now $\mathbf{x}_1 = 2\mathbf{c}_0$, as was already seen, and the higher directions $\mathbf{x}_j$ ($j \geq 2$) can be similarly expressed***; in fact,**** from the known formula

$$X_j \mu_i = \sum_{k=1}^j \frac{\partial H_{k-1}}{\partial q(0)} (2\mu_0)^{j-k} X_1 \mu_i,$$

it develops that

$$\mathbf{x}_j(\omega) = X_j \int_0^1 \omega = \sum_{k=1}^j \frac{\partial H_{k-1}}{\partial q(0)} 2^{j-k+1} \sum_{i=1}^{\infty} \frac{\phi(\mu_i)}{D^*(\mu_i)} (\mu_i)^{j-k},$$

$D(\lambda)$ now being formed from the actual auxiliary spectrum μ_i ($i = 1, 2, \dots$) for Q and not from λ_{2i-1} ($i = 1, 2, \dots$). Fortunately, this discrepancy does not

* $D = d/dx$. $\partial/\partial q$ signifies functional gradient.

** $H_2 = \int_0^1 [(1/2) q^2 + (1/4) q'^2] dx$, for example.

*** McKean-Trubowitz [6].

**** McKean-van Moerbeke [4].

matter : $\mathbf{x}_j$ is independent of the place $\mathbf{x} \in \mathbf{J}$ corresponding to Q , so it may be assumed that

$$\mu_i = \lambda_{2i-1} \quad (i = 1, 2, \dots), \text{ whence}$$

$$\mathbf{x}_j = \sum_{k=1}^j 2^{j-k+1} \frac{\partial H_{k-1}}{\partial q(0)} \mathbf{c}_{j-k}.$$

The directions $\mathbf{x}_j$ ($j = 1, 2, \dots$) will now be related to *differentials of the second kind* on $\mathbf{K}$.

A differential of the second kind [DSK] is a form $\omega = \phi(\lambda) d\lambda/y(\lambda)$ just like a differential of the first kind with two modifications :

$$(1) \int_0^\infty |\phi(\lambda)|^2 \lambda^{3/2} d\lambda < \infty \text{ is replaced by } \int_1^\infty |\phi(\lambda)|^2 \lambda^j d\lambda < \infty$$

for some $j > -\infty$; and

(2) it is required that the real periods of ω vanish*. $\Delta^\bullet d\lambda/y$ is an example, the vanishing of its real periods being plain [$y = (\Delta^2 - 1)^{1/2}$]; in fact, *modulo* DFK, the class DSK is spanned by $\lambda^{j-1} \Delta^\bullet d\lambda/y$ ($j = 1, 2, \dots$).

Proof. Let $\omega = \phi d\lambda/y \in \text{DSK}$. (2) implies that

$$\phi(\mu_i) = 0 \text{ for some } \mu_i \in [\lambda_{2i-1}, \lambda_{2i}] \quad (i = 1, 2, \dots),$$

and it follows from (1) that ϕ has only a finite number of additional roots; in fact, if $\mu_i^\bullet$ ($i = 1, \dots, k$) are additional roots, then

$$\psi(\lambda) = \phi(\lambda) \prod_{i=1}^k (\lambda - \mu_i^\bullet)^{-1}$$

satisfies $\int_0^\infty |\psi(\lambda)|^2 \lambda^{3/2} d\lambda$ for large k , and $\psi = 0$ would follow from $\psi(\mu_i) = 0$ ($i = 1, 2, \dots$), by interpolation. $\phi(\lambda)$ is now expressed as the product of a polynomial $c_0 \lambda^n + c_1 \lambda^{n-1} + \dots + c_n$ and the function $D(\lambda)$ formed from the auxiliary spectrum μ_i ($i = 1, 2, \dots$). Now $-\Delta^\bullet(\lambda)$ is also of the form $D(\lambda)$ and, the elementary estimate $D(\lambda) = \lambda^{-1/2} \sin \sqrt{\lambda} + o(\lambda^{-1})$ ($\lambda \uparrow \infty$) implies $D + \Delta^\bullet = o(\lambda^{-1})$. Thus $\phi_1(\lambda) = \phi(\lambda) + c_0 \lambda^n \Delta^\bullet(\lambda) = o(\lambda^{n-1})$ ($\lambda \uparrow \infty$) satisfies $\int_0^\infty |\phi_1(\lambda)|^2 \lambda^{j+2} d\lambda < \infty$. $\omega_1 = \phi_1 d\lambda/y$ is now adjusted by the addition of a differential of the first kind so as to have vanishing real periods**. This makes ω_1 belong to DSK, keeping $\int_1^\infty |\phi_1(\lambda)|^2 \lambda^{j+2} d\lambda < \infty$. The proof is finished by a self-evident induction over n .

* (2) takes place for $\omega \in \text{DFK}$ only if $\omega = 0$; see McKean-Trubowitz [5].

** The numbers a_i ($i = 1, 2, \dots$) are the real periods of some $\omega \in \text{DFK}$ if and only if they vanish rapidly as $i \uparrow \infty$.

Let $\lambda_0 = 0$ for simplicity. The preceding confirms the existence of unique differentials $\omega_j \in \text{DSK}$ with*

$$\int_0^p \omega_j = (2\lambda)^{j-1} \sqrt{-\lambda} + o(1) \quad (j = 1, 2, 3, \dots)$$

on nice circles. The link between differentials of the second kind and the infinitesimal motions $X_j (j = 1, 2, \dots)$ is now provided by the period relations** :

$$\sum_{i=1}^{\infty} \mathbf{a}_i(\omega) \mathbf{b}_i(\omega_j) = 2\pi \sqrt{-1} \mathbf{x}_j(\omega) \quad (j = 1, 2, \dots)$$

for differentials $\omega \in \text{DFK}$, $\mathbf{b}_i$ being customary the imaginary period obtained by clockwise integration about $[\lambda_0, \lambda_{2i-1}]$ ($i = 1, 2, \dots$).

*Proof of the period relations***.* Let C be a nice circle centred at $\lambda_0 = 0$ of radius $\frac{1}{2}(\lambda_{2n} + \lambda_{2n+1})$ with a two-banked cut along the positive, let D be the region enclosed, and let B be its full boundary. Then, by Stokes' formula,

$$\begin{aligned} 0 &= \int_D \omega_j \wedge \omega = \int_B \left[\int_0^\lambda \omega_j \right] \omega \\ &= \int_C [(2\lambda)^{j-1} \sqrt{-\lambda} + o(1)] \omega \\ &\quad - \frac{1}{2} \sum_{i=1}^n \mathbf{a}_i(\omega) \mathbf{b}_i(\omega_j), \end{aligned}$$

as a picture easily confirms. The integral about C is now expressed by means of the function $D(\lambda)$ formed for the auxiliary spectrum $\mu_i = \lambda_{2i-1}$ ($i = 1, 2, \dots$) :

$$\begin{aligned} \int_C [(2\lambda)^{j-1} \sqrt{-\lambda} + o(1)] \frac{\lambda \phi(\lambda)}{D(\lambda)} \frac{D(\lambda)}{y(\lambda)} \frac{d\lambda}{\lambda} \\ = \int_C (2\lambda)^{j-1} \left(\mathbf{c}_0 + \frac{\mathbf{c}_1}{\lambda} + \dots \right) \left[\prod_{i=1}^{\infty} \frac{\lambda_{2i-1} - \lambda}{\lambda_{2i} - \lambda} \right]^{1/2} \frac{d\lambda}{\lambda} + o(1), \end{aligned}$$

and it develops that

$$\sum_{i=1}^{\infty} \mathbf{a}_i(\omega) \mathbf{b}_i(\omega_j) = 2\pi \sqrt{-1} \times 2^j$$

$\times$ the coefficient of λ^{-j+1} in the formal power series

$$(\mathbf{c}_0 + \mathbf{c}_1 \lambda^{-1} + \dots) \prod_{i=1}^{\infty} [(\lambda_{2i-1} - \lambda)/(\lambda_{2i} - \lambda)]^{1/2}.$$

* $\mathbf{o} = (0, 0)$. $\int_0^p \omega$ is computed in the plane cut along $[0, \infty)$. The radical $y(\lambda)$ is declared to be real and positive for $\lambda < 0$. The estimate $\sqrt{-\lambda} \Delta^*/y \sim 1$ on nice circles is required.

** The i th summand is automatically dropped if $\lambda_{2i-1} = \lambda_{2i}$ [$\mathbf{a}_i(\omega) = 0$].

*** See Siegel [9] for the classical case.

This evaluation is compared to the previous formula for $\mathbf{x}_j$; the latter may be stated as*

$$\mathbf{x}_j(\omega) = 2^j \times \text{the coefficient of } \lambda^{-j+1} \text{ in the formal power series } (\mathbf{c}_0 + \mathbf{c}_1 \lambda^{-1} + \dots) \\ \times \sum_{i=1}^{\infty} (2\lambda)^{-i} \partial H_i / \partial q(0),$$

so it suffices to identify the formal power series for the product with the sum. This may be done as in McKean-van Moerbeke [4]. The formula**

$$\sum_{i=1}^{\infty} (2\lambda)^{-i} \partial H_i / \partial q(0) = \prod_{i=1}^{\infty} \frac{1 - \mu_i / \lambda}{[(1 - \lambda_{2i-1} / \lambda)(1 - \lambda_{2i} / \lambda)]^{1/2}},$$

proved there for $g < \infty$, is still valid for $g = \infty$ if the left-hand side be construed as the development of the product on nice circles tending to ∞ . The proof is finished.

The present discussion may be summed up more geometrically by the statement that DSK may be naturally identified with the part of the tangent space of J arising from the higher fields X_j ($j = 1, 2, \dots$).

5. Periodicity and units

The period relation

$$\sum_{i=1}^{\infty} \mathbf{a}_i(\omega_i) \mathbf{b}_i(\omega_i) = 2\pi \sqrt{-1} \mathbf{x}_j$$

shows that the flow $\exp(tX_j)$ is of period T if and only if $\mathbf{b}_i(\omega_j) \equiv 0 \pmod{2\pi \sqrt{-1}/T}$ ($i = 1, 2, \dots$), which is to say that the function*** $e_j(\mathbf{p}) = \exp(T \int_{\mathbf{p}} \omega_j)$ is single-valued on $\mathbf{K}$, the real periods of ω_j being trivial; in particular, for $= 1$, $\int_{\mathbf{p}} \omega_1 = \int_{\mathbf{p}} \Delta^{\circ} \mathbf{y}^{-1} d\lambda = \cosh^{-1} \Delta(\lambda)$,

$$\text{and } e_1(\mathbf{p}) = \Delta(\lambda) \pm [\Delta^2(\lambda) - 1]^{1/2} = \Delta + \mathbf{y},$$

is the multiplier of Q , i.e., if $\mathbf{p}$ and $\mathbf{p}'$ are the two points of $\mathbf{K}$ covering λ , then $e_1(\mathbf{p})$ and $e_1(\mathbf{p}')$ are the two eigenvalues of the monodromy matrix of Q . The nice thing is that e_1 is a unit**** of the function field of $\mathbf{K}$:

$$e_1(\mathbf{p}) e_1(\mathbf{p}') = (\Delta + \mathbf{y})(\Delta - \mathbf{y}) = \Delta^2 - \mathbf{y}^2 = 1,$$

and that the existence (single-valuedness) of this unit is equivalent to the periodicity of Q , as will be plain. e_1 is the unit advertised in the introduction; clearly

* Q is chosen to achieve the auxiliary spectrum $\mu_i = \lambda_{2i-1}$ ($i = 1, 2, \dots$).

** $\lambda_0 = 0$.

*** $o = (0, 0)$.

**** Let $y_0(\lambda)$ be the desingularisation of $y(\lambda)$. $e(\mathbf{p}) = a(\lambda) + y_0(\lambda) b(\lambda)$ is a unit of the function field of $\mathbf{K}$ if a and b are (transcendental) integral and $e(\mathbf{p}) e(\mathbf{p}') = a^2 - y_0^2 b^2 = 1$,

$e_1(\mathbf{p}) = \exp [\sqrt{-\lambda} + o(1)]$ on nice circles tending to ∞ and is uniquely specified by this development. e_j is also a unit if it is single-valued:

$$\mathbf{y}(\mathbf{p}') = -\mathbf{y}(\mathbf{p}),$$

so $\int_{\mathbf{p}}^{\mathbf{p}'} \omega_j = -\int_{\mathbf{p}'}^{\mathbf{p}} \omega_j$ and $e_j(\mathbf{p}) e_j(\mathbf{p}') = 1$.

The general statement is as follows: *The function field of $\mathbf{K}$ contains a unit $e(\mathbf{p})$ satisfying the reality condition that $d \lg e(\mathbf{p})$ be real on the lacunae $[\lambda_{2i-1}, \lambda_{2i}]$ ($i = 1, 2, \dots$) and enjoying, on nice circles, a development of the form:*

$$e(\mathbf{p}) = \exp \left[\sum_{j=1}^m c_j (2\lambda)^{j-1} \sqrt{-\lambda} + o(1) \right],$$

with real c_j ($j = 1, \dots, m$) if and only if the infinitesimal motion $X = \sum_{j=1}^m c_j X_j$

induces a flow of period 1; the unit is uniquely specified by its development.

Proof. The uniqueness is plain: in fact, the ratio $f(\mathbf{p}) = a(\lambda) + \mathbf{y}(\lambda) b(\lambda)$ of two such units is bounded on nice circles, so $2a(\lambda) = f(\mathbf{p}) + f(\mathbf{p}')$ and $2\mathbf{y}(\lambda) b(\lambda) = f(\mathbf{p}) - f(\mathbf{p}')$ are similarly bounded and consequently constant, and from $f(\mathbf{p}) f(\mathbf{p}') = a^2 - \mathbf{y}^2 b^2 = 1$, it develops that $a = \pm 1$, $b = 0$, as required. Now let the flow be periodic. This implies the vanishing, *modulo* $2\pi\sqrt{-1}$ of the imaginary periods of $\omega = \sum_{j=1}^m c_j \omega_j \in \text{DSK}$ and the single-valuedness of $e(\mathbf{p}) = \exp \left(\int_{\mathbf{p}} \omega \right)$. The sheet map $\mathbf{p} \rightarrow \mathbf{p}'$ flips the signature of the integral, so e is a unit, and the reality of $\omega = d \lg e(\mathbf{p})$ on the lacunae is plain from the fact that ω is contained in the real span of $\lambda^{j-1} \Delta^*(\lambda) d\lambda/\mathbf{y}(\lambda)$ ($j = 1, \dots, m$) and DFK; in short, the periodicity implies existence of the unit. Now let a unit e of the required kind exist. $e(\mathbf{p})$ cannot vanish [$e(\mathbf{p}) e(\mathbf{p}') = 1$], so all the periods of $\omega = d \lg e(\mathbf{p})$ vanish, *modulo* $2\pi\sqrt{-1}$; also, ω is real on the lacunae, so its real periods vanish entirely. The development of $e(\mathbf{p})$ at ∞ now confirms that $\omega \in \text{DSK}$. The proof is finished.

6. Toda units

Let $\mathbf{K}$ be a non-singular hyperelliptic curve of genus $g < \infty$ with $2g + 2$ finite branch points $\lambda_0 < \lambda_1 < \dots < \lambda_{2g+1}$. Let $\mathbf{y}(\lambda)$ be the radical $[(\lambda - \lambda_0) \dots (\lambda - \lambda_{2g+1})]^{1/2}$ and $\mathbf{p} = [\lambda, \mathbf{y}(\lambda)]$ the general point of $\mathbf{K}$, and let $\mathbf{a}_i, \mathbf{b}_i$ ($i = 1, \dots, g$) be the customary homology basis. $\mathbf{K}$ is a *Toda curve* if any of the following equivalent conditions hold:

(1) $\mathbf{K}$ has a unit* $e(\mathbf{p}) = a(\lambda) + \mathbf{y}(\lambda) b(\lambda)$ with rational integral a and b satisfying the reality condition that $d \lg e(\mathbf{p})$ be real on the lacunae $[\lambda_{2i-1}, \lambda_{2i}]$ ($i = 1, \dots, g$).**

* $e(\mathbf{p}) e(\mathbf{p}') = 1$, $\mathbf{p} \rightarrow \mathbf{p}'$ being the sheet map.

** The finite-genus curves of § 2 do not admit such units, the corresponding radicals $\mathbf{y}(\mathbf{p})$ having an odd number of finite branch points,

$$(2) \sum_{i=1}^g m_i a_i(\omega) = n \int_{\mathbf{p}_\infty}^{\mathbf{p}'_\infty} \omega$$

for* $\omega \in \text{DFK}$, $n = \deg e$, and integral m_i ($i = 1, \dots, g$), $\mathbf{p}_\infty$ and $\mathbf{p}'_\infty$ being the two points of $\mathbf{K}$ covering $\lambda = \infty$.**

(3) $\mathbf{K}$ is the desingularisation of a curve $e^2 - 2\Delta e + 1 = 0$ defined by the vanishing of

$$\det \begin{bmatrix} b_{1-\lambda} & a_1 & 0 & \dots 0 & a_n/e \\ a_1 & b_{2-\lambda} & a_2 & \dots 0 & 0 \\ \vdots & & & & \vdots \\ a_n e & 0 & 0 & \dots a_{n-1} & b_{n-\lambda} \end{bmatrix} = c(e + e^{-1} - 2\Delta)$$

with real

$$a_i, b_i (i = 1, \dots, n)$$

$$\text{and} \quad c = \prod_{i=1}^n a_i;$$

evidently, $\Delta(\lambda) = d_0 \lambda^n + \dots + d_n$ with real d_i ($i = 0, \dots, n$). $e = \Delta \pm [\Delta^2 - 1]^{1/2}$ is the unit of (1); also, $\lambda_0 < \dots < \lambda_{2g+1}$ is the simple periodic and anti-periodic spectrum of

$$Q = \begin{bmatrix} b_1 & a_1 & 0 & \dots & 0 & a_n \\ a_1 & b_2 & a_2 & \dots & 0 & 0 \\ \vdots & & & & & \vdots \\ a_n & 0 & 0 & \dots & a_{n-1} & b_n \end{bmatrix},$$

and it turns out that the class of all such matrices with the same spectra can be identified as the real part $\mathbf{J}$ of the Jacobian of $\mathbf{K}$ via the composite map***

$$Q \rightarrow \mathbf{p}_1 + \dots + \mathbf{p}_g \rightarrow \sum_{i=1}^g \int_{\mathbf{o}_i}^{\mathbf{p}_i} \omega, \quad \omega \in \text{DFK},$$

in which the divisor $\mathbf{p}_i = [\mu_i, \mathbf{y}(\mu_i)]$ ($i = 1, \dots, g$) is formed from the points of the spectrum of Q , with its last row and column deleted, such as fall in the lacunae $[\lambda_{2i-1}, \lambda_{2i}]$ ($i = 1, \dots, g$). The analogy with § 2 will be plain; see Moerbeke [7] for background and additional information about this and the sequel.

Now the isospectral flows of Q can be expressed as**** $Q^\bullet = [Q, P^+ - P^-]$, in which P is a real polynomial in Q without constant term, of degree $\leq g$. The

* DFK denotes the class of differentials of the first kind.

** $\mathbf{p}_\infty$ and/or $\mathbf{p}'_\infty$ is a division point of the curve.

*** $\mathbf{o}_i = [\lambda_{2i-1}, 0]$ ($i = 1, \dots, g$).

**** The ij entry of $P^+ - P^-$ is $+P_{ij}[-P_{ij}]$ above [below] the diagonal, and 0 on the diagonal, except that $-P_{1n}[+P_{n1}]$ appears in the upper right-hand [lower left-hand] corner.

purpose of this article is to prove that if $p(\lambda)$ is the corresponding polynomial of λ , then the flow is of period 1 if and only if $\mathbf{K}$ admits a unit $e = a + \mathbf{y} b$, satisfying the reality condition that $d \lg e(\mathbf{p})$ be real on the lacunae $[\lambda_{2i-1}, \lambda_{2i}]$ ($i = 1, \dots, g$) and having, at $\mathbf{p}_\infty$, a transcendental singularity of the form $\exp[p(\lambda) + o(1)]$.

Proof. The flow is expressed on $\mathbf{J}$ by*

$$\sum_{i=1}^g \int_{\mathbf{p}_i(0)}^{\mathbf{p}_i(t)} \omega = t \times 2 \times \text{the residue of } p\omega \text{ at } \mathbf{p}_\infty$$

for $\omega \in \text{DFK}$, $\mathbf{p}_1 + \dots + \mathbf{p}_g$ being the divisor of $\mathbf{K}$ corresponding to Q , and the periodicity of the flow means that $2 \times (\text{res } p\omega \text{ at } \mathbf{p}_\infty)$ is a period of $\mathbf{J}$:

$$2 \times (\text{res } p\omega \text{ at } \mathbf{p}_\infty) = \sum_{i=1}^g n_i a_i(\omega).$$

Now pick $\omega_1 \in \text{DSK}^{**}$ so as to have*** $\int_{\mathbf{o}} \omega_1 = p(\lambda) + o(1)$ near $\mathbf{p}_\infty$ and let ω_1 be pole-free elsewhere. Then, by the period relations for differentials of the first and second kinds,

$$2\pi\sqrt{-1} \times (\text{res } p\omega \text{ at } \mathbf{p}_\infty) = -(1/2) \sum_{i=1}^g a_i(\omega) b_i(\omega_1)$$

for $\omega \in \text{DFK}$, from which it appears that the flow is of period 1 if and only if $b_i(\omega_1) \equiv 0 \pmod{2\pi\sqrt{-1}}$ ($i = 1, \dots, g$). Now the real periods of $\omega_1 \in \text{DSK}$ vanish automatically, so $e(\mathbf{p}) = \exp(\int_{\mathbf{o}} \omega_1)$ is single-valued on $\mathbf{K}$; also, e is a unit since the sheet map $\mathbf{p} \rightarrow \mathbf{p}'$ flips the signature of the integral; it is of the required form $\exp[p(\lambda) + o(1)]$ at $\mathbf{p}_\infty$; and it satisfies the reality condition that $d \lg e(\mathbf{p}) = \omega_1$ be real on the lacunae $[\lambda_{2i-1}, \lambda_{2i}]$ ($i = 1, \dots, g$) since it belongs to the real span of $\lambda^{j-1} d\lambda/\mathbf{y}(\lambda)$ ($j = 1, 2, \dots$). Thus, the periodicity implies the existence of the unit. Now let such a unit e exist. $e(\mathbf{p}) e(\mathbf{p}') = 1$ so $e(\mathbf{p}) \neq 0$, and the periods of $\omega_1 = d \lg e(\mathbf{p})$ vanish, modulo $2\pi\sqrt{-1}$; in fact, its real periods vanish by the reality of ω_1 on the lacunae, and the development $\omega_1 \sim \mathbf{p}'(\lambda) d\lambda$ at $\mathbf{p}_\infty$ confirms $\omega_1 \in \text{DSK}$. The insertion of ω_1 into the periodicity relations now yields the periodicity of the flow.

References

- [1] Borg G 1945 Eine Umkehrung der Sturm-Liouvilleschen Eigenwertaufgabe. *Acta Math.* **78** 1-96
- [2] Krichever L M 1976 Algebraic-geometric construction of Zakharov-Shabat equations and their solutions. *Sov. Math. Dokl.* **17** 394-397
- [3] Marcenko V and Ostrovskii 1975 A characterisation of the spectrum of Hill's operator. *Sov. Math. Dokl.* **16** 761-765; *Mat. Sbornik* **97** 540-606

* See Moerbeke [7] and Moerbeke-Mumford [8].

** DSK denote the class of differentials of the second kind with vanishing real periods.

*** $\mathbf{o}$ is any branch point of $\mathbf{K}$.

- [4] McKean H and van Moerbeke P 1975 The spectrum of Hill's equation. *Inventiones Math.* **30** 217-274
- [5] McKean H and Trubowitz E 1976 Hill's operator and hyper-elliptic function theory in the presence of infinitely many branch points. *Comm. Pure Appl. Math.* **29** 143-226
- [6] McKean H and Trubowitz E 1978 Hill's surfaces and their theta functions. *Bull. Am. Math. Soc.* **84** 1042-1085
- [7] Moerbeke P van 1978 The spectrum of Jacobi matrices. *Inventiones Math.* **37** 45-81
- [8] Moerbeke P van and Mumford D 1978 The spectrum of difference operators and the theory of curves. *Acta Math.* **143** 93-154
- [9] Siegel C L 1971 *Topics in complex function theory* (New York: Wiley Interscience) Vol. 2
- [10] Trubowitz E 1977 The inverse problem for periodic potentials. *Comm. Pure Appl. Math.* **30** 321-337

Courant Institute of Mathematical Sciences,
New York
NY 10012, USA

A submersion principle and its applications

By

HARISH-CHANDRA

1. Introduction

Let G be a real reductive group and π an irreducible admissible representation of G . Let Θ denote the character of π . We recall that Θ is a distribution on G defined by

$$\Theta(f) = \text{tr } \pi(f) \quad (f \in C_c^\infty(G)).$$

It is well known that Θ is a locally summable function on G which is analytic on the set G' of regular elements. Fix $\gamma_0 \in G'$ and let Γ be the Cartan subgroup of G containing γ_0 . Put

$$\Gamma' = G' \cap \Gamma \text{ and } G_\Gamma = \bigcup_{x \in G} x \Gamma' x^{-1}.$$

The mapping $(x, \gamma) \mapsto x\gamma x^{-1}$ of $G \times \Gamma'$ into G is everywhere submersive. Since Θ is invariant under all inner automorphisms of G , one proves easily that Θ defines a distribution θ on Γ' . Let $\mathcal{Z}$ be the algebra of all differential operators on G which commute with both left and right translations. Then Θ satisfies the differential equations

$$z\Theta = \chi(z)\Theta \quad (z \in \mathcal{Z}),$$

where χ is the infinitesimal character of π . We can transcribe these equations in terms of θ . In this way we obtain a system of differential equations for θ on Γ' . It turns out that this system is elliptic and therefore θ is an analytic function on Γ' . But this implies that Θ coincides with an analytic function on G_Γ .

We would like to prove a similar result in the p -adic case. Let Ω be a p -adic field and G the group of all Ω -rational points of a connected reductive Ω -group $\mathbf{G}$ [3]. Then G , with its usual topology, is a locally compact, totally disconnected, unimodular group. Let dx denote its Haar measure. We shall use the terminology of [3] without further comment.

Let π be an admissible irreducible representation of G . Then for every $f \in C_c^\infty(G)$, the operator

$$\pi(f) = \int_G f(x) \pi(x) dx$$

has finite rank. Put

$$\Theta(f) = \text{tr } \pi(f).$$

Then Θ is a distribution on G . Let G' be the set of all points $x \in G$ where $D_G(x) \neq 0$ ([3], § 15). Then G' is an open, dense subset of G whose complement has measure zero. We intend to show that Θ coincides with a locally constant function F on G' . This means that

$$\Theta(f) = \int f(x) F(x) dx,$$

for all $f \in C_c^\infty(G')$.

In case $\text{char } \Omega = 0$, this fact was first proved by Howe by making use of his Kirillov theory for p -adic groups. Moreover when $\text{char } \Omega = 0$, it is known that Θ is a locally summable function on G [4, 5]. But these methods, which make extensive use of Lie algebras and the exponential mapping, do not seem to work in characteristic p . We shall therefore construct a proof on a totally different principle.

2. The submersion principle

Let P be a parabolic subgroup of G and $x \mapsto x^*$ the projection of G on $G^* = G/P$.

Theorem 1 (the submersion principle). Fix $\gamma \in G'$. Then the mapping

$$x \mapsto (x\gamma x^{-1})^*$$

from G to G^* is everywhere submersive.

When $\text{char } \Omega = 0$, the proof is very easy. Moreover Borel assures me that this principle is actually true over an arbitrary field.

Let us verify it when $\text{char } \Omega = 0$. Put

$$\phi_\gamma: x \mapsto (x\gamma x^{-1})^*.$$

Then $\phi_\gamma(xy) = \phi_{\gamma\gamma^{-1}}(x)$ ($x, y \in G$).

Hence it is enough to prove that ϕ_γ is submersive at $x = 1$.

Fix a split component A of P and let $P = MN$ be the corresponding Levi decomposition of P . Let $(\bar{P}, A)$ denote the p -pair opposite to (P, A) . Then $\bar{P} = M\bar{N}$. We denote the Lie algebras of $G, P, M, N, \bar{N}$ by $\mathfrak{g}, \mathfrak{p}, \mathfrak{m}, \mathfrak{n}, \bar{\mathfrak{n}}$ respectively. Then

$$\mathfrak{p} = \mathfrak{m} + \mathfrak{n}, \quad \mathfrak{g} = \bar{\mathfrak{n}} + \mathfrak{m} + \mathfrak{n},$$

and we have to verify that

$$(\text{Ad } (\gamma^{-1}) - 1) \mathfrak{g} + \mathfrak{p} = \mathfrak{g}.$$

Fix a symmetric, nondegenerate, G -invariant bilinear form B on $\mathfrak{g}$ with values in Ω . Since G is reductive, this is possible. Let Γ be the Cartan subgroup of G determined by γ and $\mathfrak{t}$ the Lie algebra of Γ . Then $\mathfrak{t} = \ker (\text{Ad } (\gamma) - 1)$. For

any linear subspace $\mathfrak{q}$ of $\mathfrak{g}$, let $\mathfrak{q}^\perp$ denote the space of all $Y \in \mathfrak{g}$ such that $B(X, Y) = 0$ for all $X \in \mathfrak{q}$. Then

$$\mathfrak{p}^\perp = \mathfrak{n}, ((\text{Ad}(\gamma^{-1}) - 1)\mathfrak{g})^\perp = \ker (\text{Ad}(\gamma) - 1) = \mathfrak{r}.$$

Therefore $((\text{Ad}(\gamma^{-1}) - 1)\mathfrak{g} + \mathfrak{p})^\perp \subset \mathfrak{r} \cap \mathfrak{n} = \{0\}$,

and this implies the desired result.

3. The function $f_{a, \gamma}$

Fix γ and P as in theorem 1. Then the mapping*

$$(x, p) \mapsto {}^x\gamma \cdot p \quad (x \in G, p \in P)$$

of $G \times P$ into G is everywhere submersive. Hence ([2], p. 49) there exists a unique linear mapping

$$a \mapsto f_{a, \gamma} \quad (a \in C_c^\infty(G \times P))$$

from $C_c^\infty(G \times P)$ to $C_c^\infty(G)$ such that

$$\int_{G \times P} a(x : p) F({}^x\gamma \cdot p) dx d_1 p = \int_G f_{a, \gamma}(x) F(x) dx,$$

for all $F \in C_c^\infty(G)$. Here $d_1 p$ is the left Haar measure on P .

Lemma 1. Fix $a \in C_c^\infty(G \times P)$. Then the mapping

$$y \mapsto f_{a, \gamma} \quad (y \in G')$$

from G' to $C_c^\infty(G)$ is locally constant.

The mapping

$$(y, x, p) \mapsto (y, {}^x\gamma \cdot p)$$

from $G' \times G \times P$ to $G' \times G$ is submersive. Hence there exists a unique linear mapping

$$\beta \mapsto \phi_\beta \quad (\beta \in C_c^\infty(G' \times G \times P))$$

from $C_c^\infty(G' \times G \times P)$ to $C_c^\infty(G' \times G)$ such that

$$\int \beta(y : x : p) \Phi(y : {}^x\gamma \cdot p) dy dx d_1 p = \int \phi_\beta(y : x) \Phi(y : x) dy dx$$

for all $\Phi \in C_c^\infty(G' \times G)$. Now let

$$\Phi(y : x) = \lambda(y) F(x),$$

where $\lambda \in C_c^\infty(G')$ and $F \in C_c^\infty(G)$. Then

$$\int_G \lambda(y) dy \int_{G \times P} \beta(y : x : p) F({}^x\gamma \cdot p) dx d_1 p = \int_G \lambda(y) dy \int_G \phi_\beta(y : x) F(x) dx.$$

* We write ${}^x\gamma = x\gamma x^{-1}$ for $x, \gamma \in G$.

This being true for all λ , we conclude that

$$\int_{G \times P} \beta(y : x : p) F({}^a y \cdot p) dx d_1 p = \int \phi_\beta(y : x) F(x) dx$$

for all $y \in G'$ and $F \in C_c^\infty(G)$.

Now fix $y_0 \in G'$ and put

$$\beta(y : x : p) = \mu(y) \alpha(x : p)$$

where $\alpha \in C_c^\infty(G \times P)$, $\mu \in C_c^\infty(G')$ and $\mu(y_0) = 1$. Let G_0 be a neighbourhood of y_0 in G' such that $\mu = 1$ on G_0 . Then $\beta(y : x : p) = \alpha(x : p)$ for $y \in G_0$ and therefore

$$\begin{aligned} \int_{G \times P} \beta(y : x : p) F({}^a y \cdot p) dx d_1 p &= \int \alpha(x : p) F({}^a y \cdot p) dx d_1 p \\ &= \int f_{\alpha, y}(x) F(x) dx. \end{aligned}$$

Hence $\int f_{\alpha, y}(x) F(x) dx = \int \phi_\beta(y : x) F(x) dx$

for $y \in G_0$ and $F \in C_c^\infty(G)$. This shows that

$$f_{\alpha, y}(x) = \phi_\beta(y : x) \quad (y \in G_0, x \in G).$$

Since $\phi_\beta \in C_c^\infty(G' \times G)$, our assertion is now obvious.

4. The operator $T_\bullet$

Fix a minimal p -pair (P, A) in G ($P = MN$) and let K be an open compact subgroup of G of Bruhat-Tits ([2], p. 16) corresponding to A . Let π be an admissible representation of G on V . We recall that $\text{End}^0 V$ is defined to be the subspace of all $T \in \text{End } V$ such that the mappings

$$x \mapsto \pi(x) T, \quad x \mapsto T \pi(x),$$

of G into $\text{End } V$ are both smooth.

Let dk denote the normalised Haar measure on K .

Theorem 2. Let π be an admissible representation of G on V such that V is a finite G -module under π . For $x \in G'$, define

$$T_\bullet = \int_K \pi(k x k^{-1}) dk.$$

Then $T_\bullet \in \text{End}^0 V$ and $x \mapsto T_\bullet$ is a smooth mapping from G' to $\text{End}^0 V$.

Let K_0 be an open and normal subgroup of K and V_0 the subspace of all vectors in V , which are left fixed by K_0 . Then $\dim V_0 < \infty$. By choosing K_0 sufficiently small, we may assume that V is spanned by elements of the form $\pi(x) v$ ($x \in G, v \in V_0$).

Let M^+ denote the set of all $m \in M$ such that $\langle \alpha, H_M(m) \rangle \geq 0$ for every root α of (P, A) (see [3], § 7) for the definition of H_M). Then $G = KM^+ \cdot K$. Put $A^+ = A \cap M^+$. Since M/A is compact, we conclude that $M^+ \subset CA^+$ where C is a compact subset of M . Hence it is clear that we can choose an open compact subgroup P_0 of P such that $m^{-1}P_0m \subset K_0$ for all $m \in M^+$.

Let α denote the characteristic function of $K \times P_0$ and put $f_y = f_{\alpha, y}$ ($y \in G'$) in the notation of lemma 1. Then $y \mapsto f_y$ is a smooth mapping from G' to $C_0^\infty(G)$ and

$$\int_{K \times P_0} F(ky \cdot p) dk dp = \int_G f_y(x) F(x) dx,$$

for all locally summable functions F on G . (Here dp is the normalised Haar measure on P_0 .) From this we deduce immediately that

$$\int_{K \times P_0} \pi(ky \cdot p) dk dp = \int_G f_y(x) \pi(x) dx = \pi(j_y).$$

Let V_y be the smallest K -invariant subspace of V containing $\pi(j_y)V$. Since $f_y \in C_0^\infty(G)$, it is clear that $\dim V_y < \infty$.

Since K_0 is normal in K , V_0 is stable under $\pi(K)$. Therefore since $G = KM^+K$ V is spanned by $\pi(KM^+)V_0$. Fix $k \in K$, $m \in M^+$ and $v \in V_0$. Then

$$T_y \pi(km)v = \pi(k) T_y \pi(m)v \quad (y \in G').$$

$$\text{But} \quad \pi(f_y) \pi(m)v = T_y \int_{P_0} \pi(pm)v dp = T_y \pi(m) \int_{P_0} \pi(m^{-1}pm)v dp = T_y \pi(m)v$$

since $m^{-1}P_0m \subset K_0$. Therefore

$$T_y \pi(km)v = \pi(k) \pi(f_y) \pi(m)v \in V_y.$$

This shows that $T_y V \subset V_y$. Since T_y commutes with $\pi(k)$ ($k \in K$), $\dim V_y < \infty$ and π is admissible, it is now clear that $T_y \in \text{End}^0 V$.

Fix k, m, v as above. Since the mapping $y \mapsto f_y$ is smooth, it follows from the result obtained above that the mapping

$$y \mapsto T_y \pi(km)v = \pi(k) \pi(f_y) \pi(m)v \quad (y \in G'),$$

from G' to V is also smooth. The second statement of the theorem is now obvious if we recall that V is spanned by $\pi(KM^+)V_0$.

Corollary. Let Θ denote the character of π . Then Θ coincides on G' with the locally constant function

$$x \mapsto \text{tr } T_x \quad (x \in G').$$

$$\text{Put } f^0(x) = \int_K f(kxk^{-1}) dk \quad (x \in G),$$

for $f \in C_0^\infty(G)$. Then

$$\Theta(f) = \Theta(f^0) = \text{tr } \pi(f^0).$$

But
$$\pi(f^0) = \int_G f(x) T_x dx$$

for $f \in C_c^\infty(G')$. Hence

$$\Theta(f) = \int_G f(x) \operatorname{tr} T_x dx \quad (f \in C_c^\infty(G')).$$

5. Some applications

Let Γ be a Cartan subgroup of G . For $\gamma \in \Gamma' = \Gamma \cap G'$ and $f \in C_c^\infty(G)$, define

$$F_\Gamma(\gamma) = |D(\gamma)|^{\frac{1}{2}} \int_{G/A_\Gamma} f(x\gamma x^{-1}) dx^*$$

where $D = D_G$, A_Γ is the split component of Γ and dx^* is an invariant measure on G/A_Γ .

Theorem 3. Let K_0 be an open compact subgroup of G . Given $\gamma_0 \in \Gamma'$, we can choose a neighbourhood ω of γ_0 in Γ' such that F_Γ is constant on ω for every $f \in C_c(G/K_0)$.

This result had been proved by Howe some years ago in the case $\operatorname{char} \Omega = 0$.

Without loss of generality, we may assume that K_0 is a normal subgroup of K . Let us now use the notation of §4. Then

$$\int_{K \times P_0} F({}^k\gamma \cdot p) dk dp = \int_G f_\gamma(x) F(x) dx,$$

for $\gamma \in \Gamma'$ and $F \in C_c^\infty(G)$. Fix $f \in C_c(G/K_0)$ and put

$$g(x) = \int_K f(kxk^{-1}) dk \quad (x \in G).$$

Since K_0 is normal in K , $g \in C_c(G/K_0)$. Fix $m \in M^+$ and let $F(x) = g(m^{-1}xm)$ ($x \in G$). Then

$$\int_{K \times P_0} g(m^{-1} \cdot {}^k\gamma \cdot pm) dk dp = \int f_\gamma(x) g(m^{-1}xm) dx.$$

But $pm = m \cdot m^{-1}pm \in mK_0$. Hence

$$\int_K g(m^{-1} \cdot {}^k\gamma \cdot m) dk = \int f_\gamma(x) g(m^{-1}xm) dx \quad (\gamma \in \Gamma').$$

Fix a neighbourhood Γ_0 of γ_0 in Γ' such that $f_\gamma = f_{\gamma_0}$ for $\gamma \in \Gamma_0$. This is possible by lemma 1. Then

$$\int_K g(m^{-1} \cdot {}^k\gamma \cdot m) dk = \int_K g(m^{-1} \cdot {}^k\gamma_0 \cdot m) dk \quad (\gamma \in \Gamma_0).$$

By standard arguments, the proof of theorem 3 is reduced to the case when Γ is elliptic. Then $A_\Gamma = Z$ where Z is the maximal split torus lying in the centre of G . We know from the work of Bruhat and Tits that

$$K \backslash G/KZ \simeq M^+/^0MZ.$$

$$\begin{aligned} \text{Therefore } F_f(\gamma) &= |D(\gamma)|_{\mathfrak{p}}^{\frac{1}{2}} \int_{G/\mathbb{Z}} f(x\gamma x^{-1}) dx^* \\ &= |D(\gamma)|_{\mathfrak{p}}^{\frac{1}{2}} \sum_{m \in M^+ \cap \mathbb{Z}} \mu(m) \int_K g(m^{-1} \cdot {}^k\gamma \cdot m) dk, \end{aligned}$$

where $\mu(m)$ is the Haar measure of KmK . Let ω be a neighbourhood of γ_0 in Γ_0 such that $|D(\gamma)|_{\mathfrak{p}}$ is constant for $\gamma \in \omega$. Then

$$F_f(\gamma) = F_f(\gamma_0) \quad (\gamma \in \omega).$$

Since ω is independent of $f \in C_c(G/K_0)$, the theorem is proved.

Theorem 3 makes it possible to prove Lemma 13 of ([3], § 16) without any restriction on char Ω .

I believe it is important, from the point of view of harmonic analysis, to obtain an analogue of theorem 5 of ([1], p. 32). We give below a result which may be regarded as a step in this direction.

Take $K_0 = K$ and define $f_\gamma (\gamma \in \Gamma')$ as in § 4. Then $f_\gamma \geq 0$. Put

$$\beta(\gamma) = \sup_x f_\gamma(x) \quad (\gamma \in \Gamma')$$

and define Ξ as in ([3], § 14).

Theorem 4. Let ω be a compact subset of Γ . Then we can choose a positive number c such that

$$\int_K \Xi(m^{-1} \cdot {}^k\gamma \cdot m) dk \leq c\beta(\gamma) \Xi(m)^2,$$

for all $m \in M^+$ and $\gamma \in \omega' = \omega \cap \Gamma'$.

It is obvious that

$$\text{Supp } f_\gamma \subset \bigcup_{{}^k\gamma \in K} ({}^k\gamma \cdot P_0). \quad (\gamma \in \Gamma').$$

Therefore $\text{Supp } f_\gamma \subset {}^k\omega \cdot P_0 = C$ (say)

for $\gamma \in \omega'$. Since C is compact, we can choose a finite number of elements y_i ($1 \leq i \leq r$) in G such that

$$\text{Supp } f_\gamma \subset \bigcup_{1 \leq i \leq r} y_i K$$

for all $\gamma \in \omega'$.

Now fix $\gamma \in \omega'$, $m \in M^+$ and put

$$F(x) = \Xi(m^{-1}xm) \quad (x \in G)$$

in the relation

$$\int_{K \times P_0} F({}^k\gamma \cdot p) dk dp = \int_G f_\gamma(x) F(x) dx.$$

Observe that

$$F({}^k\gamma \cdot p) = \Xi(m^{-1} \cdot {}^k\gamma \cdot pm) = \Xi(m^{-1} \cdot {}^k\gamma \cdot m) \quad (p \in P_0)$$

since $m^{-1}pm \in K$. Therefore

$$\begin{aligned} \int_K \Xi(m^{-1} \cdot {}^k\gamma \cdot m) dk &\leq \beta(\gamma) \sum_i \int_K \Xi(m^{-1} y_i km) dk \\ &= \beta(\gamma) \sum_i \Xi(m^{-1} y_i) \Xi(m) \end{aligned}$$

from the identity

$$\int_K \Xi(xky) dk = \Xi(x) \Xi(y) \quad (x, y \in G).$$

We can choose a number $c_1 \geq 1$ such that

$$\Xi(xy_i) \leq c_1 \Xi(x) \quad (1 \leq i \leq r)$$

for all $x \in G$. Put $c = rc_1$ and observe that $\Xi(x^{-1}) = \Xi(x)$. Then we get

$$\int_K \Xi(m^{-1} \cdot {}^k\gamma \cdot m) \leq c\beta(\gamma) \Xi(m)^2$$

and this proves our assertion.

One would like to verify that

$$\sup_{\gamma \in \omega'} |D(\gamma)|_p^{\frac{1}{2}} \beta(\gamma) < \infty.$$

I believe this to be true but do not have a proof.

References

- [1] Harish-Chandra 1966 Discrete series for semisimple Lie groups, II. *Acta Math.* **116** 1–111.
- [2] Harish-Chandra 1970 *Harmonic analysis on reductive p -adic groups*, *Lecture Notes in Math.* (Berlin and New York: Springer-Verlag) Vol. 162.
- [3] Harish-Chandra 1973 Harmonic analysis on reductive p -adic groups, in *Harmonic analysis on homogeneous spaces* (Providence: Am. Math. Soc.), pp. 167–192.
- [4] Harish-Chandra 1977 The characters of reductive p -adic groups, in *Contributions to algebra* (New York: Academic Press), pp. 175–182.
- [5] Harish-Chandra 1978 Admissible invariant distributions on reductive p -adic groups, Lie theories and their applications, *Queen's papers in pure and applied mathematics*, No. 48 (1978), Queen's University, Kingston, Ontario, pp. 281–347.

Institute for Advanced Study,
Princeton, New Jersey 08540, USA

Geometric construction of cohomology for arithmetic groups I

By

JOHN J MILLSON* and M S RAGHUNATHAN*

1. Introduction

In this paper we give a geometric construction of cohomology classes for uniform arithmetic subgroups of Lie groups and prove that our classes are not the restrictions of continuous classes from the Lie group. We apply our theory to obtain the following results:

(1) a proof that every unitary representation with non-zero cohomology, see Borel [2], occurs in $L^2(\Gamma \backslash SO(n, 1))$ for a suitable uniform Γ , thereby proving for the group $SO(n, 1)$ a conjecture which seems to be widely believed, though the group $SO(n, 1)$ represents the only case where it has been proved. In terms of differential geometry we show that the n dimensional manifolds of constant negative curvature constructed at the end of Borel [4] admit finite covers with all Betti numbers strictly between 0 and n not equal to zero—and hence arbitrarily large by Borel [2];

(2) a proof that the recent vanishing theorem of Borel-Wallach [5], Casselman-Schmid [6] and Zuckerman [13] is the best possible for the orthogonal groups $SO(p, q)$ for $p \geq q$ and q even;

(3) the construction of algebraic cycles in compact quotients of the unit ball in $\mathbb{C}^n$ which are dual to linear subspaces of CP^n but correspond to non-trivial automorphic representations. We construct similar examples for some other bounded symmetric domains, namely, those associated to $SU(p, q)$ and $SO(n, 2)$.

We remark that the identification of the local factors in the automorphic representations of (3) appears to be an important problem.

We now describe our construction.

If Γ is a torsion-free discrete subgroup of a semi-simple Lie group G having no compact simple factors and K is a maximal compact subgroup of G then Γ acts freely on the Riemannian symmetric space $X = G/K$ and consequently the cohomology of the quotient space $\Gamma \backslash X$ coincides with the cohomology of Γ . We deduce the existence of new cohomology for Γ by constructing non-bounding cycles in $\Gamma \backslash X$.

* Both authors were partially supported by NSF Grant # MCS-76-10435.

These cycles will be the projections under $\pi : X \rightarrow \Gamma \backslash X$ of the sub-symmetric spaces of X which are fixed point sets of involutive isometries of X which normalise Γ . It is important to observe that the class of sub-symmetric spaces of X which are fixed point sets of involutions is a proper sub-class of the totally geodesic sub-symmetric spaces of X . To try to prove that for suitable Γ the projection of an arbitrary sub-symmetric space of X gives rise to a non-bounding cycle would be unsuccessful as it would contradict the above vanishing theorem. In fact, the class of sub-symmetric spaces of X which arise as the fixed point sets of isometric involutions is characterised as the class of sub-symmetric spaces admitting complements. By this we mean the following: Suppose 0 is the origin of X and X_1 a totally geodesic sub-symmetric space of X passing through 0. Let Φ_1 be the tangent space to X_1 at 0 and Φ_2 the orthogonal complement of Φ_1 in Φ , which is identified with the tangent space of X at 0. Let X_2 be the image of Φ_2 under the Riemannian exponential map $\exp : T(X, 0) \rightarrow X$. We say X_2 is a complement to X_1 if X_2 is a sub-symmetric space of X . In case X_1 is the fixed point set of an involutive isometry σ_1 this is always the case, for if ι is the geodesic reflection at 0 then X_2 is the fixed point set of the involutive isometry $\sigma_2 = \iota\sigma_1$. Projecting X_1 and X_2 down to $\Gamma \backslash X$ we obtain two complementary dimensional closed cycles. By passing to a subgroup Γ' of finite index in Γ we obtain a new projection $\Pi' : X \rightarrow \Gamma' \backslash X$. We try to find Γ' so that the cycles $\Pi'(X_1)$ and $\Pi'(X_2)$ intersect transversally with *all* intersection multiplicities positive. In this case the intersection product of the two cycles is non-zero and consequently neither is a boundary. One of our most important theorems, theorem 2.1, then asserts that if one can find such a sub-group then there exists a further subgroup Γ'' of finite index in Γ so that $\Pi''(X_1)$ is not dual to an invariant form, where $\Pi'' : X \rightarrow \Gamma'' \backslash X$ is the projection.

We would like to express our appreciation to W. Dwyer for the observation at the beginning of the proof of theorem 2.1.

1. Orthogonal decompositions of symmetric spaces

We begin by establishing some notation which will be used in the rest of the paper. X will denote a symmetric space of non-compact type which is assumed to be connected. G will denote the connected component of the identity of the symmetry group of X , G will be assumed to be linear and semi-simple. The symmetric space X will be equipped with an origin 0. The isotropy at 0 will be denoted by K ; hence, X will be isometric to G/K equipped with a metric in the usual way. $\mathfrak{G}$ will denote the Lie algebra of G , $\mathfrak{k}$ the Lie algebra of K and Φ the orthogonal complement of $\mathfrak{k}$ in $\mathfrak{G}$ relative to the Cartan-Killing form. ι will denote the geodesic reflection at 0.

The key notion in this paper is the following. Let X_1 and X_2 be two totally geodesic sub-symmetric spaces of X whose intersection is precisely 0. Let Φ_1 denote the tangent space to X_1 at 0 and Φ_2 the tangent space to X_2 at 0. We say that the pair $\{X_1, X_2\}$ is an orthogonal decomposition of X if

(i) Φ_1 and Φ_2 are perpendicular

(ii) $\Phi = \Phi_1 \oplus \Phi_2$.

We say that two orthogonal decompositions $\{X_1, X_2\}$ and $\{X'_1, X'_2\}$ are equivalent if there is an isometry ϕ of X so that $\phi(X_1) = X'_1$ and $\phi(X_2) = X'_2$.

Remark : Since X is homogeneous there is no point in considering decompositions $\{X_1, X_2\}$ so that $X_1 \cap X_2 = \{x\}$ where $x \neq 0$.

Now we see orthogonal decompositions may be characterised infinitesimally by the following proposition.

Proposition 1.1.

Let $\eta \subseteq \Phi$ be a sub-vector space. Then the following are equivalent.

- (1) $Y = \exp \eta$ is a totally geodesic sub-symmetric space
- (2) η is a Lie triple system, that is if $u, v, w \in \eta$, then $[[u, v], w] \in \eta$.

Proof : The proof may be found in Helgason [7], theorem 7-2, page 189.

Thus the infinitesimal equivalent of an orthogonal decomposition of X is an orthogonal decomposition $\Phi = \Phi_1 \oplus \Phi_2$ with Φ_1 and Φ_2 both Lie triple systems.

We now show how all such arise.

Theorem 1.1

There is a one to one correspondence between equivalence classes of orthogonal decompositions of X and conjugacy classes of involutive isometries of X . This correspondence is obtained as follows: Given an involutive isometry σ_1 of X fixing 0 put $\sigma_2 = \sigma_1 \iota$ (note that σ_1 and ι commute so that σ_2 is an involution). Then denoting by X_1 the fixed point set of σ_1 and by X_2 the fixed point set of σ_2 we find that $\{X_1, X_2\}$ is an orthogonal decomposition.

Proof : Suppose that we are given σ_1 as above. Then defining σ_2, X_1, X_2 as above it is clear that $\{X_1, X_2\}$ is an orthogonal decomposition of X .

To prove the converse consider the infinitesimal decomposition $\Phi = \Phi_1 \oplus \Phi_2$. Then Φ_1 and Φ_2 are both triple systems. We define a linear map T on Φ by $T|_{\Phi_1} = +I$, $T|_{\Phi_2} = -I$ and T preserves Φ_1 and Φ_2 .

If we can show that T preserves the curvature tensor R at 0 then by Cartan's theorem T will be the derivative of a global isometry σ_1 evaluated at 0. Thus we must show that for all $x, y, z, w \in \Phi$ we have

$$\langle R_{Tx, Ty} Tw, Tz \rangle = \langle R_{x, y} w, z \rangle.$$

But using the well-known formula for the curvature tensor of a symmetric space, see Helgason [7], theorem 4.2, page 180.

$$R_{x, y} z = -[[x, y], z],$$

decomposing x, y, z, w into Φ_1 and Φ_2 components, using that Φ_1 and Φ_2 are triple systems and that the Cartan-Killing form is invariant the result is immediately verified and theorem 1.1 is proved.

We will often denote the decomposition $\{X_1, X_2\}$ of X by σ_1, σ_2 . Here σ_1, σ_2 are a pair of commuting involutive isometries fixing 0 and satisfying

$$\sigma_1 \sigma_2 = \iota.$$

Lemma 1.1

Two orthogonal decompositions (σ_1, σ_2) , (σ'_1, σ'_2) are equivalent if and only if there exists an isometry ϕ of X fixing 0 and satisfying

$$\phi\sigma_1\phi^{-1} = \sigma'_1,$$

$$\phi\sigma_2\phi^{-1} = \sigma'_2.$$

Proof: The sufficiency is clear. Suppose then that there is an isometry ϕ of X satisfying

$$\phi(X_1) = X'_1,$$

$$\phi(X_2) = X'_2,$$

where X'_1 and X'_2 are the fixed point sets of σ'_1 and σ'_2 . ϕ map $X_1 \cap X_2$ to $X'_1 \cap X'_2$ hence $\phi(0) = 0$. Denoting the tangent spaces of X'_1 and X'_2 at 0 by Φ'_1 and Φ'_2 respectively we have :

$$d\phi(\Phi_1) = \Phi'_1,$$

$$d\phi(\Phi_2) = \Phi'_2,$$

$$d(\phi\sigma_1\phi^{-1})|_{\Phi'_1} = +1,$$

$$d(\phi\sigma_2\phi^{-1})|_{\Phi'_2} = -1.$$

But $\phi\sigma_1\phi^{-1}$ and σ'_1 both leave 0 fixed and have the same derivative at 0. Since they are isometries they must be equal. A similar argument establishes that $\phi\sigma_2\phi^{-1} = \sigma'_2$. This proves lemma 1.1.

We define two subgroups of G

$$G_1 = \{g \in G \mid \sigma_1 g \sigma_1 = g\},$$

$$G_2 = \{g \in G \mid \sigma_2 g \sigma_2 = g\}.$$

We will denote by $\overline{G}_1$ and $\overline{G}_2$ the subgroups of G_1 and G_2 which preserve the orientation of X_1 and X_2 respectively. It is also convenient to define involutions

$$\Sigma_1(g) = \sigma_1 g \sigma_1,$$

$$\Sigma_2(g) = \sigma_2 g \sigma_2,$$

$$\theta(g) = \iota g \iota.$$

Then $\theta = \Sigma_1 \Sigma_2$ and is the Cartan involution associated to K . We will sometimes denote the decomposition by $\{\Sigma_1, \Sigma_2\}$.

Lemma 1.2

$\overline{G}_1$ acts transitively on X_1 and $\overline{G}_2$ acts transitively on X_2 .

Proof: $P_1 = \exp \Phi_1$ is contained in $\overline{G}_1$ and acts transitively on X_1 . Similarly $P_2 = \exp \Phi_2$ acts transitively on X_2 . (Here $\exp$ is the exponential $g \rightarrow G$).

Remark. In our applications G_1 and G_2 will be replaced by smaller groups H_1 and H_2 . These groups will be defined in Chapter 4. For the moment we let H_1 and H_2 be subgroups of G_1 and G_2 respectively which act transitively on X_1 and X_2 respectively. We let H_1^0 and H_2^0 denote the identity components of H_1 and H_2 respectively.

Lemma 1.3

Put $\hat{K} = G_1 \cap G_2$. Then $\hat{K} = G_1 \cap K = G_2 \cap K$ and is a maximal compact subgroup of G_1 and G_2 . $\hat{K}$ is reductive but it is frequently not semi-simple.

Proof: $K \cap G_1$ is a maximal compact subgroup of G_1 because it is the isotropy of 0. We have $G_1 \cap G_2 \subseteq K \cap G_1$ because $G_1 \cap G_2$ must leave 0 fixed and consequently is contained in K . But suppose $k \in G_1 \cap K$. Then $\theta(k) = k$ because $k \in K$ and $\Sigma_1(k) = k$ because $k \in G_1$, hence $\Sigma_2(k) = k$, hence $k \in G_2$. This proves lemma 1.3.

Remark. Putting lemmas 1.2 and 1.3 together we see

$$X_1 = G_1/\hat{K}, \quad X_2 = G_2/\hat{K}.$$

Proposition 1.2

Let G be a connected simply-connected semi-simple real Lie group and let σ be an involutive automorphism. Then the fixed point set of σ is a connected subgroup.

Proof: In Helgason [7], theorem 7.2, page 272, the theorem is proved in case G is compact. Now assume G is not compact. σ leaves invariant some maximal compact subgroup K . Then we obtain a σ -equivariant decomposition $G = KP$. But the fixed point set of σ in P is clearly connected because if a point $x \in P$ is fixed, since 0 is also fixed, the minimal geodesic joining 0 to x is pointwise fixed. By the theorem referred to in the first line of this proof the fixed point set of σ in K is connected (K is simple-connected because G is) hence the fixed point set of σ in G is the product of two connected sets. This proves proposition 1.2.

2. Discrete uniform subgroups compatible with an orthogonal decomposition

Definition

We say a discrete subgroup $\Gamma \subseteq G$ is compatible with a decomposition $\{\sigma_1, \sigma_2\}$ if σ_1 and σ_2 both preserve Γ ; that is,

$$\sigma_1 \Gamma \sigma_1 \subseteq \Gamma,$$

$$\sigma_2 \Gamma \sigma_2 \subseteq \Gamma.$$

We define subgroups Γ_1 and Γ_2 of Γ by

$$\Gamma_1 = \{\gamma \in \Gamma \mid \sigma_1 \gamma \sigma_1 = \gamma\},$$

$$\Gamma_2 = \{\gamma \in \Gamma \mid \sigma_2 \gamma \sigma_2 = \gamma\}.$$

Remark : σ_1 and σ_2 induce involutions of $\Gamma \backslash X$ which we also denote by σ_1 and σ_2 .

Proposition 2.1

Suppose that $\Gamma \subseteq G$ is uniform; then Γ_1 is uniform in G_1 and Γ_2 is uniform in G_2 .

Proof : We will need two lemmas.

Lemma 2.1 (Selberg)

Let G be a locally compact group and let Γ be a discrete subgroup. If $\Gamma \backslash G$ is compact then for all $\gamma \in \Gamma$.

$Z(\gamma) \cap \Gamma \backslash Z(\gamma)$ is compact where $Z(\gamma)$ denotes the centraliser of γ in G .

Proof : Let $\gamma \in \Gamma$. Consider the map $k : G \rightarrow G$ given by $k(g) = g\gamma g^{-1}$. Then k is continuous. Since $k(\Gamma) \subset \Gamma$, $k(\Gamma)$ is closed and accordingly $k^{-1}(k(\Gamma))$ is closed, that is, $\Gamma Z(\gamma)$ is closed in G . Therefore $\Gamma \backslash \Gamma Z(\gamma)$ is a closed subset of the compact space G and is compact. Hence $Z(\gamma) \cap \Gamma \backslash Z(\gamma)$ is compact.

Remark : The proof of this lemma is taken from Mostow [11], page 62.

The next lemma is hardly more than an observation.

Lemma 2.2

Let G be a topological group and Γ a discrete subgroup. Let G_0 be a subgroup of G of finite index in G . Then Γ is uniform in G if and only if $\Gamma \cap G_0$ is uniform in G_0 .

Proof : It is clear that if $\Gamma \cap G_0$ is uniform in G_0 then Γ is uniform in G . To prove the converse it is sufficient to show that $\Gamma \cap G_0 \backslash G_0$ is closed in $\Gamma \backslash G$. This is equivalent to showing ΓG_0 is closed in G . But ΓG_0 is just a union of a finite number of copies of G_0 hence a finite union of closed sets proving lemma 2.2.

Now we are ready to prove proposition 2.1. Consider $G = \text{Aut } G$. We have a natural projection $\Pi : G \rightarrow \text{Ad } G$ and $\text{Ad } G$ is a subgroup of finite index of $\text{Aut } G$. $\Pi(\Gamma)$ is a uniform discrete subgroup of $\text{Ad } G$ hence of $\text{Aut } G$. Consider the subgroup Γ' of $\text{Aut } G$ generated by $\Pi(\Gamma)$ together with σ_1 and σ_2 . Clearly Γ' is uniform but it is also easily seen to be discrete as σ_1 and σ_2 normalise Γ . Thus we may apply lemma 2.1 to conclude that the centraliser of σ_1 in Γ' is uniform in the centraliser of σ_1 in $\text{Aut } G$ which we denote $Z(\sigma_1)$. But by lemma 2.2, $Z(\sigma_1) \cap \Gamma' \cap \text{Ad } G$ is uniform in $Z(\sigma_1) \cap \text{Ad } G$ since $Z(\sigma_1) \cap \text{Ad } G$ is of finite index in $Z(\sigma_1)$ (here we use that $\text{Ad } G$ is normal in $\text{Aut } G$). From this we see that

the inverse image of $Z(\sigma_1) \cap \Gamma' \cap \text{Ad } G$ is uniform in the inverse image of $Z(\sigma_1) \cap \text{Ad } G$ which we denote $(G_1)_*$. But G_1 has finite index in $(G_1)_*$ and

$$\Gamma_1 = G_1 \cap \Pi^{-1}(Z(\sigma_1) \cap \Gamma' \cap \text{Ad } G).$$

We apply lemma 2.2 and obtain that Γ_1 is uniform in G_1 and of course the same proof applies to Γ_2 in G_2 .

It is not generally the case that $Y_1 = \Gamma_1 \backslash X_1$ and $Y_2 = \Gamma_2 \backslash X_2$ will be the images of X_1 and X_2 under the quotient projection $\Pi : X \rightarrow \Gamma \backslash X$. However, somewhat surprisingly, this is always the case if Γ acts freely on X .

Proposition 2.2 (H Jaffee)

With the above notation if Γ operates freely on X :

$$\Pi(X_1) = \Gamma_1 \backslash X_1,$$

$$\Pi(X_2) = \Gamma_2 \backslash X_2.$$

Proof: Suppose there exist $x_2 \in X_2$, $x'_2 \in X$, $\gamma \in \Gamma$ so that $\gamma x_2 = x'_2$. Then let $v = \gamma \sigma_2 \gamma^{-1} \sigma_2$. Then since $\sigma_2 \gamma^{-1} \sigma_2 \in \Gamma$, $v \in \Gamma$. But $v x'_2 = x'_2$ hence $v = 1$. From this it follows $\sigma_2 \gamma^{-1} \sigma_2 = \gamma^{-1}$ and accordingly $\gamma \in \Gamma_2$. A similar argument establishes $\Pi(X_1) = \Gamma_1 \backslash X_1$.

We receive then (assuming Γ is compatible with $\{\sigma_1, \sigma_2\}$ and acts freely) a compact locally symmetric space Y containing two totally geodesic complementary dimensional totally geodesic submanifolds Y_1 and Y_2 . It will be very important to us that Y_1 , Y_2 and Y are orientable.

Lemma 2.2

(i) Y is orientable.

(ii) If $\Gamma_1 \subseteq \tilde{G}_1$ then Y_1 is orientable.

Proof: (i) is obvious since Γ is assumed to be contained in G which is connected. (ii) follows immediately from proposition 2.2. We note then if G is simply-connected then from proposition 1.2 it follows that Y_1 and Y_2 are orientable.

In fact we may generalise this last remark considerably.

Proposition 2.3

Suppose G has a finite linear covering group H with the property that σ_1 lifts to an involution $\tilde{\sigma}_1$ of H with a connected centraliser H_1 . Then there exists an arithmetic subgroup Γ' of Γ so that $\Gamma'_1 \subseteq \tilde{G}_1$ where Γ'_1 is the centraliser of σ_1 in Γ' .

Proof: Let Δ be the kernel of the projection $\Pi : H \rightarrow G$. The finite central extension $\Delta \rightarrow H \rightarrow G$ induces an extension $\Delta \rightarrow \tilde{\Gamma} \rightarrow \Gamma$. We claim that this extension splits over a subgroup Γ' of Γ of finite index. Indeed, since H is linear we can find a congruence subgroup $\tilde{\Gamma}'$ of $\tilde{\Gamma}$ containing no elements of

finite order and normalised by $\tilde{\sigma}_1$ hence $\Delta \cap \tilde{\Gamma}' = \{1\}$. Put $\Gamma' = \Pi(\tilde{\Gamma}')$. Clearly Γ' is an arithmetic subgroup of Γ . Noting that $\tilde{\Pi} : \Gamma' \rightarrow \Gamma'$ is an isomorphism we see that $\Delta \rightarrow \Pi^{-1}(\Gamma') \rightarrow \Gamma'$ is split. Now we claim that Π maps the centraliser of $\tilde{\sigma}_1$ in $\tilde{\Gamma}'$ onto the centraliser of σ_1 in Π' . Since the first centraliser is contained in the connected group H_1 , this will complete the proof of the proposition. Suppose $\gamma \in \Gamma'$ satisfies $\sigma_1 \gamma \sigma_1 = \gamma$. Let $\tilde{\gamma}$ be the element in $\tilde{\Gamma}'$ lying over γ . Then $\Pi(\tilde{\sigma}_1 \tilde{\gamma} \tilde{\sigma}_1^{-1}) = 1$ and since $\tilde{\sigma}_1 \tilde{\gamma} \tilde{\sigma}_1^{-1} \in \tilde{\Gamma}'$ we must have $\tilde{\sigma}_1 \tilde{\gamma} \tilde{\sigma}_1^{-1} = \tilde{\gamma}$ and the proposition is proved.

Corollary. Suppose the universal cover $\tilde{G}$ of G is linear. Then there exists an arithmetic subgroup Γ' of Γ so that $\tilde{\Gamma}' \subseteq G_1$ and $\Gamma'_2 \subseteq \overline{G}_2$.

Remark : In case G is the group of analytic automorphisms of a bounded symmetric domain then $\tilde{G}$ is not linear. However if σ_1 is complex linear (and consequently $\sigma_2 = i\sigma_1$ is also complex linear) then G_1 is a group of analytic automorphisms of the sub-domain $G_1/\hat{K}$ and consequently every element of G_1 is orientation preserving hence $G_1 = \overline{G}_1$ (and $G_2 = \overline{G}_2$).

Lemma 2.3

There is a one to one correspondence between $Y_1 \cap Y_2$ and the set of $\Gamma_1 \times \Gamma_2$ equivalence classes of triples

$$T = \{(x_1, x_2) \in X_1 \times X_2 \times \Gamma \mid \gamma x_1 = x_2\}.$$

$\Gamma_1 \times \Gamma_2$ acts on this set by

$$(\gamma_1, \gamma_2) \circ (x_1, x_2, \gamma) = (\gamma_1 x_1, \gamma_2 x_2, \gamma_2 \gamma \gamma_1^{-1}).$$

Proof : Let (x_1, x_2, γ) be an equivalence class representative then $\gamma x_1 = x_2$ hence $\Pi(x_1) = \Pi(x_2) \in Y_1 \cap Y_2$. Now suppose $y \in Y_1 \cap Y_2$, then there exists $x_1 \in X_1$, $x_2 \in X_2$ so that $\Pi(x_1) = y = \Pi(x_2)$ hence there exists $\gamma \in \Gamma$ so that $\gamma x_1 = x_2$ hence y corresponds to the triple (x_1, x_2, γ) . Now let $x'_1 \in X_1$, $x'_2 \in X_2$ be two other choices so that $\Pi(x'_1) = y = \Pi(x'_2)$. Then there exists γ' so that $\gamma' x'_1 = x'_2$. By proposition 2.2 there exists γ_1, γ_2 so that $\gamma_1 x'_1 = x_1$, $\gamma_2 x'_2 = x_2$ hence $\gamma_2^{-1} \gamma \gamma_1(x'_1) = \gamma'(x'_1)$ hence $\gamma_2 \gamma' \gamma_1^{-1} = \gamma$ establishing lemma 2.3.

Proposition 2.3

Let $(x_1, x_2, \gamma) \in T$, then the connected component of $Y_1 \cap Y_2$ passing through $y = \Pi(x_1)$ is just the image of $\gamma(X_1) \cap X_2$ under Π . Let $\sigma'_1 = \gamma \sigma_1 \gamma^{-1}$. Then the intersection component containing $\Pi(x_1)$ is a locally symmetric space $\Lambda \backslash A/B$ where A is the centraliser of σ'_1 in G_2 , B is the centraliser of σ'_1 in $\hat{K}$ and Λ is the centraliser of σ'_1 in Γ_2 .

Proof : Let y' be in the connected component of $Y_1 \cap Y_2$ containing y . Then according to the correspondence of lemma 2.3 there exist $x_1 \in X_1$, $x'_2 \in X_2$ and

$\gamma' \in \Gamma$ so that $\gamma' x'_1 = x'_2$ and $\Pi(x'_1) = \Pi(x'_2) = \gamma'$. But there is a path $a \in Y_1 \cap Y_2$ joining y to y' , hence under the correspondence of lemma 2.3 there must exist paths $a_1(t) \in X_1$, $a_2(t) \in X_2$ and $\gamma(t)$ in Γ so that $\Pi \circ a_1 = \Pi \circ a_2 = a$. Since Γ is discrete we must have $\gamma(t) = \gamma$ all t ; hence $\gamma x'_1 = x'_2$ proving that the connected component of $Y_1 \cap Y_2$ containing y is contained in $\Pi(\gamma(X_1) \cap X_2)$.

To prove the converse it is enough to show that $\gamma(X_1) \cap X_2$ is connected. To this end let $x_2, x'_2 \in \gamma(X_1) \cap X_2$, then there exists $x_1, x'_1 \in X_1$ with $\gamma(x_1) = x_2$, $\gamma(x'_1) = x'_2$. But then the minimising geodesic joining x_1 to x'_1 is sent to the minimising geodesic joining x_2 to x'_2 and accordingly the latter geodesic is entirely contained in the intersection.

The above description of the intersection component containing $\Pi(x_1)$ follows immediately from the Jaffee lemma.

Remark : Every isolated intersection has multiplicity $+1$ or -1 since $Y_1 \cap Y_2$ is totally geodesic.

We see that the connected components of $Y_1 \cap Y_2$ are indexed by the Γ_2, Γ_1 double cosets of those γ which appear in the third component of the triples in T . Choose a set of representatives for the Γ_2, Γ_1 double cosets of these elements of Γ , $\gamma_1, \gamma_2, \dots, \gamma_r$ and put $I(\Gamma) = \{\gamma_1, \gamma_2, \dots, \gamma_r : \gamma_1 = 1\}$. (We will sometimes write I for $I(\Gamma)$.)

Lemma 2.4

There is a one to one correspondence between I and the $\Gamma_2 \times \Gamma_1$ orbits of $\Gamma \cap H_2^0 KH_1^0$.

Proof : Suppose $\gamma \in I$, then there exists $x_1 \in X_1$ and $x_2 \in X_2$ so that $\gamma x_1 = x_2$. Choose $h_1 \in H_1^0$ so that $h_1 \cdot 0 = x_1$ and $h_2 \in H_2^0$ so that $h_2 \cdot 0 = x_2$, then $\delta(\gamma) = h_2^{-1} \gamma h_1 \in K$ hence $\gamma \in H_2^0 KH_1^0$.

Conversely, given $\gamma \in H_2^0 KH_1^0$ we may choose h_2, h_1 so that $h_2^{-1} \gamma h_1 \in K$ hence $\gamma h_1 \cdot 0 = h_2 \cdot 0$ and $x_2 = h_2 \cdot 0$ we have $\gamma x_1 = x_2$. This proves lemma 2.4.

We now establish a formula for the multiplicity of an intersection corresponding to $\gamma \in I$. Given $\gamma \in I$ we may associate $\delta(\gamma) \in K$ as above. In what follows we will write δ for $\delta(\gamma)$.

Proposition 2.3

Let $e_1, e_2, \dots, e_k$ be a frame for $T(X_1, 0)$ and $e_{k+1}, e_{k+2}, \dots, e_n$ be a frame for $T(X_2, 0)$. Consider the real number $\epsilon(\gamma)$ defined by

$$\epsilon(\gamma) = \langle \delta_*(e_1 \wedge e_2 \wedge \dots \wedge e_k), e_1 \wedge e_2 \wedge \dots \wedge e_k \rangle$$

where $\langle, \rangle$ is the Riemannian inner product on k -vectors at x . $\epsilon(\gamma)$ satisfies then $\delta_*(e_1 \wedge e_2 \wedge \dots \wedge e_k) \wedge e_{k+1} \wedge \dots \wedge e_n = \epsilon(\gamma) e_1 \wedge e_2 \wedge \dots \wedge e_n$. Then

- (a) $\epsilon(\gamma) = 0$ if and only if the intersection is degenerate,
- (b) $\epsilon(\gamma)$ is positive if and only if the intersection multiplicity is $+1$,
- (c) $\epsilon(\gamma)$ is negative if and only if the intersection multiplicity is -1 .

Notation : If $f: X \rightarrow X$ is a smooth map and $v_1 \wedge v_2 \wedge \dots \wedge v_p$ is a p -vector tangent to x then $f_*(v_1 \wedge v_2 \wedge \dots \wedge v_p) = f_*v_1 \wedge f_*v_2 \wedge \dots \wedge f_*v_p$ is a p -vector tangent to $f(x)$.

Proof : $\omega = (h_1)_*(e_1 \wedge e_2 \wedge \dots \wedge e_k)$ is the positive orientation of X_1 at x_1 and $\tau = (h_2)_*(e_{k+1} \wedge e_{k+2} \wedge \dots \wedge e_n)$ is the positive orientation of X_2 at x_2 . Thus the intersection will be positive, negative or degenerate according to whether the n vector $\gamma_*(\omega) \wedge \tau$ is a positive, negative or zero multiple of the volume element of X at x_2 . But up to a positive scalar multiple this volume element is just $(h_2)_*(e_1 \wedge e_2 \wedge \dots \wedge e_n)$; thus, we must evaluate the sign of the inner product

$$\langle \gamma_*(\omega) \wedge \tau, (h_2)_*(e_1 \wedge e_2 \wedge \dots \wedge e_n) \rangle_{x_2}.$$

But h_2 is an isometry of the Riemannian metric on n -vectors; hence, the above inner product is the same as

$$\begin{aligned} & \langle (h_2^{-1})_*(\gamma_*(\omega) \wedge \tau), e_1 \wedge e_2 \wedge \dots \wedge e_n \rangle_0 \\ &= \langle (h_2^{-1})_*((\gamma h_1)_*(e_1 \wedge \dots \wedge e_k) \wedge (h_2)_*(e_{k+1} \wedge \dots \wedge e_n)), e_1 \wedge \dots \wedge e_n \rangle_0 \\ &= \langle (h_2 \gamma h_1)_*(e_1 \wedge \dots \wedge e_k) \wedge e_{k+1} \wedge \dots \wedge e_n, e_1 \wedge \dots \wedge e_n \rangle_0 \\ &= \langle (\delta)_*(e_1 \wedge \dots \wedge e_k) \wedge e_{k+1} \wedge \dots \wedge e_n, e_1 \wedge \dots \wedge e_n \rangle_0 \\ &= \epsilon(\gamma). \end{aligned}$$

This proves proposition 2.3.

We warn the reader that in the degenerate case $\epsilon(\gamma) = 0$ is not necessarily the intersection multiplicity.

Note that the value of $\epsilon(\gamma)$ does not depend on the choice of h_1 and h_2 because any two choices of h_1 differ by multiplication by an element of $\bar{K} \cap \bar{G}_1$ (which does not change $\epsilon(\gamma)$ as $K \cap G_1$ preserves Φ_1 and Φ_2 and their orientations) and the same is true for h_2 . Now we prove a lemma which is critical for us.

Lemma 2.5

If $\gamma, \gamma' \in I$ are in the same $\bar{G}_2, \bar{G}_1$ double coset in G then $\epsilon(\gamma) = \epsilon(\gamma')$.

Proof : Suppose $\gamma' = h_2^* \gamma h_1^*$ and we choose h_2', h_1' so that $h_2'^{-1} \gamma' h_1' = \delta(\gamma') \in K$. Then putting $h_2 = h_2'^{-1} h_2'$ and $h_1 = h_1^* h_1'$ we have $h_2^{-1} \gamma h_1 = h_2'^{-1} \gamma h_1^* h_1' = \delta(\gamma') \in K$.

Case 1. The intersection corresponding to γ is nondegenerate. In this case there is a unique x_1 (up to action by Γ_1) and a unique x_2 (up to action by Γ_2) so that $\gamma x_1 = x_2$, for if there were x_1' and x_2' inequivalent under Γ_1 and x_2 and x_2' corresponding then the minimal geodesic joining x_1 to x_1' in X_1 would be carried onto the minimal geodesic joining x_2 and x_2' in X_2 and consequently the intersection would be degenerate. Now we have $h_2^{-1} \gamma h_1 \circ 0 = 0$ consequently the choice of $h_2 = h_2'^{-1} h_2'$ and $h_1 = h_1^* h_1'$ is a correct choice to calculate $\epsilon(\gamma)$ as in proposition 2.3. But for this choice we have $\delta(\gamma) = \delta(\gamma')$ hence $\epsilon(\gamma) = \epsilon(\gamma')$ in the non-degenerate case.

Case 2. The intersection corresponding to γ is degenerate. In this case there is an entire geodesic in X_1 which is carried into a geodesic in X_2 . But there γ' will carry h_1^{*-1} of this geodesic into h_2^* of the image geodesic in X_2 . Thus $\epsilon(\gamma) = 0$ if and only if $\epsilon(\gamma') = 0$. This proves the lemma.

Corollary

Suppose γ is in the trivial $\bar{G}_2, \bar{G}_1$ double coset; that is, $\gamma \in \bar{G}_2 \bar{G}_1$; then $\epsilon(\gamma) = +1$.

We shall use the following notation in the rest of this paper. If G is an algebraic group over a number field E and F is a field containing E then $G(F)$ will denote the group of F -rational points of G . More generally, if R is any ring, then $G(R)$ or $(G)_R$ will denote the group of R rational points of G , assuming this makes sense. Also if O_E is the integers in E we will often denote $G(O_E)$ by Φ .

We assume now that we are given an orthogonal decomposition $\{\sigma_1, \sigma_2\}$ of G/K and a compatible uniform torsion-free arithmetic subgroup Γ of G . We assume $\Gamma \subseteq SL(N, O_E)$ where E is a finite algebraic extension of $\mathbb{Q}$. For any ideal a in O_E we define the congruence subgroups.

$$\Gamma(a) = \{\gamma \in \Gamma : \gamma \equiv 1 \pmod{a}\},$$

$$\Gamma_1(a) = \{\gamma \in \Gamma_1 : \gamma \equiv 1 \pmod{a}\},$$

$$\Gamma_2(a) = \{\gamma \in \Gamma_2 : \gamma \equiv 1 \pmod{a}\}.$$

We define the manifolds

$$Y(a) = \Gamma(a) \backslash X,$$

$$Y_1(a) = \Gamma_1(a) \backslash X_1,$$

$$Y_2(a) = \Gamma_2(a) \backslash X_2.$$

We also define finite groups

$$\psi = \Gamma/\Gamma(a),$$

$$\psi_1 = \Gamma_1/\Gamma_1(a),$$

$$\psi_2 = \Gamma_2/\Gamma_2(a).$$

We now prove one of our main theorems.

Theorem 2.1

Assume that all intersections of Y_1 and Y_2 are of positive multiplicity. Then there exists a prime p in E so that the cohomology class corresponding to $Y_1(p)$ under Poincaré duality cannot be represented by an invariant form.

Proof: We begin with the observation due to W. Dwyer that if the cohomology class dual to $Y_1(p)$ could be represented by an invariant form then the homology class carried by $Y_1(p)$ is invariant under the deck transformations of the covering $\Pi : Y(p) \rightarrow Y$. This follows immediately from the naturality of the Poincaré duality map combined with the fact that the deck transformations preserve the orientation class of $Y(p)$.

The following lemmas then complete the proof of the theorem.

Lemma 2.6

There exists $\mathfrak{p}$ a prime ideal in E and $\gamma \in \Gamma$ so that $\sigma_2 \gamma \sigma_2 \gamma^{-1}$ is not congruent to an element of Γ_1 modulo $\mathfrak{p}$.

Proof: Suppose first that we can choose $\gamma \in \Gamma$ so that $\mu = \sigma_2 \gamma \sigma_2 \gamma^{-1} \notin \Gamma_1$. Then we claim that there exists $\mathfrak{p}$ so that $\sigma_2 \gamma \sigma_2 \gamma^{-1}$ is not congruent modulo $\mathfrak{p}$ to any element of Γ_1 . Indeed choose a rational representation ρ of G on a vector space V so that G_1 is the isotropy of the line through $v \in V$. Then $\rho(\mu)v$ is not on the line through v because $\mu \notin G_1$. We may then complete $\{v, \rho(\mu)v\}$ to a rational basis for V . Now let $\mathfrak{p}$ be any prime so that $\mathfrak{p}$ does not divide either the numerator or the denominator of this discriminant.

We now prove that there exists $\gamma \in \Gamma$ so that $\sigma_2 \gamma \sigma_2 \gamma^{-1} \notin \Gamma_1$. Suppose no such γ exists. We define a map $\phi: G_{\mathbb{C}} \rightarrow G_{\mathbb{C}}$ by $\phi(g) = \sigma_2 g \sigma_2 g^{-1}$. Since $\phi(\Gamma) \subseteq \Gamma_1$ and Γ is Zariski dense in $G_{\mathbb{C}}$ we must have $\phi(G_{\mathbb{C}}) \subseteq (G_1)_{\mathbb{C}}$ and hence $d\phi(g) \subseteq \mathfrak{g}_1$ where for $x \in \mathfrak{g}$ we have $d\phi(x) = \sigma_2 x \sigma_2 - x$. Now we note that $\text{Ad } \sigma_2$ induces the identity on $\mathfrak{k}_1$ since $\text{Ad } \sigma_2$ is the Cartan involution of $\mathfrak{g}_1$. But if $v_1 \in \Phi_1$, $v_2 \in \Phi_2$ and $x = [v_1, v_2]$ then $d\phi(x) = -2x \notin \mathfrak{g}_1$. With this last statement we obtain a contradiction and the lemma is proved.

Lemma 2.7

Let γ be as in lemma 2.5 and let η be the deck transformation of the covering $\Pi: Y(\mathfrak{p}) \rightarrow Y$ represented by γ . Then $\eta Y_1(\mathfrak{p})$ and $Y_1(\mathfrak{p})$ are not homologous.

Proof: All intersections of $Y_1(\mathfrak{p})$ and $Y_2(\mathfrak{p})$ are positive since every intersection of $Y_1(\mathfrak{p})$ and $Y_2(\mathfrak{p})$ lies over an intersection of Y_1 and Y_2 . Also $\eta Y_1(\mathfrak{p})$ projects to Y_1 , which does not bound, hence $\eta Y_1(\mathfrak{p})$ does not bound. Thus to show that $Y_1(\mathfrak{p})$ and $\eta Y_1(\mathfrak{p})$ are homologically independent it is sufficient to show that the intersection number of $\eta Y_1(\mathfrak{p})$ and $Y_2(\mathfrak{p})$ is zero. We show in fact that $\sigma_2(\eta Y_1(\mathfrak{p})) \cap \eta Y_1(\mathfrak{p}) = \emptyset$. Since any point in $\eta Y_1(\mathfrak{p}) \cap Y_2(\mathfrak{p})$ is fixed by σ_2 this will show that this latter set is also empty.

Let $a: \psi \rightarrow \psi$ denote the involution of ψ induced by σ_2 . Thus $\sigma(\eta) = \overline{\sigma_2 \gamma \sigma_2}$ where $\overline{\quad}$ denotes reduction modulo $\mathfrak{p}$. Let v denote the class of η in ψ/ψ_1 and β the involution of ψ/ψ_1 induced by a . By the choice of γ we have $\beta(v) \neq v$. But we claim $\sigma_2(\eta Y_1(\mathfrak{p})) = a(\eta) Y_1(\mathfrak{p})$. Indeed let $x \in Y_1(\mathfrak{p})$, then $\sigma_2 \eta x = \sigma_2 \eta \sigma_2 \sigma_2 x = a(\eta) \sigma_2 x \in a(\eta) Y_1(\mathfrak{p})$ since σ_2 stabilises $Y_1(\mathfrak{p})$. But $a(\eta) \neq \eta$ and hence $\eta Y_1(\mathfrak{p})$ and $a(\eta) Y_1(\mathfrak{p})$ are disjoint and lemma 2.7 is proved. With this the proof of theorem 2.1 is completed.

3. The Galois cocycle associated to an intersection

We assume in this chapter that we are given an orthogonal decomposition $\{X_1, X_2\}$ of the symmetric space X and a compatible uniform torsion-free arithmetic subgroup $\Gamma \subset G$. We assume $\Gamma \subset GL(N, E)$ where E is a totally real number field and that the groups G , H_1 and H_2 of Chapter 1 are the real rational points of groups M , M_1, M_2 defined over E .

Let $\mathfrak{p}$ be a prime ideal in E . We have defined the congruence subgroup $\Gamma(\mathfrak{p})$ of Γ of level $\mathfrak{p}$. We have also defined a finite set, $I(\Gamma(\mathfrak{p}))$, in chapter 2. We will abbreviate $I(\Gamma(\mathfrak{p}))$ to $I(\mathfrak{p})$ and, as before, we will abbreviate $I(\Gamma)$ to I . It is false that $I(\mathfrak{p}) \subseteq I$. However, let us define δI , $\delta I(\mathfrak{p})$, again following notation established in chapter 2 (above proposition 2.3) by:

$$\begin{aligned}\delta I &= \{\delta \gamma_2, \dots, \delta \gamma_r \mid \gamma_i \in I\}, \\ \delta I(\mathfrak{p}) &= \{\delta \gamma_1, \dots, \delta \gamma_s \mid \gamma_i \in I(\mathfrak{p})\}.\end{aligned}$$

Lemma 3.1

$$\delta I(\mathfrak{p}) \subseteq \delta I.$$

Proof: Let $\gamma \in I(\mathfrak{p})$. Then there exists $\gamma_2 \in \Gamma_2$, but not necessarily in $\Gamma_2(\mathfrak{p})$, and $\gamma_1 \in \Gamma_1$ so that $\gamma_2 \gamma \gamma_1 \in I$. But by lemma 3.1 $\delta(\gamma_2 \gamma \gamma_1) = \delta(\gamma)$ hence $\delta(\gamma) \in \delta I$.

Our strategy in the rest of the paper is to try to pass to a congruence subgroup of Γ , $\Gamma(\mathfrak{p})$ so that $\delta I(\mathfrak{p})$ is contained in the connected component of the normaliser of $\hat{K}$. We now begin our programme of trying to decrease δI .

Theorem 3.1

For all but finitely many primes $\mathfrak{p}$ we have $\gamma \in I(\mathfrak{p})$ implies $\gamma \in (H_2)_{\mathbb{C}}(H_1)_{\mathbb{C}}$ —recall $(H_2)_{\mathbb{C}}$ is the complexification of H_2 and $(H_1)_{\mathbb{C}}$ is the complexification of H_1 .

Proof: The theorem will follow immediately from the next two lemmas.

Lemma 3.2

The double cosets $(H_2)_{\mathbb{C}} \gamma_i (H_1)_{\mathbb{C}}$ ($\gamma_i \in I$) are closed in $(G)_{\mathbb{C}}$.

Proof: By a theorem of Birkes [1] it is sufficient to show that the double cosets $H_2 \gamma_i H_1$ are closed in G . Note that $H_2 \gamma_i H_1$ is closed in G if and only if $H_2 \gamma_i H_1 \gamma_i^{-1}$ is. Now we will prove $H_2 H_1$ is closed in G and the proof will be valid for $H_2 \gamma_i H_1 \gamma_i^{-1}$. Let $\mathbf{v}$ be an integral vector in V (a representation space for M) so that M_2 is precisely the isotropy of $\mathbf{v}$. Consider the orbit of $\mathbf{v}$ in $V_{\mathbb{R}}$. It is enough to prove that the image of H_1 is closed in $G\mathbf{v}$. But $\Gamma_1 \mathbf{v}$ is discrete, hence closed in $V_{\mathbb{R}}$ by Borel [3], 7.13 and $H_1 = \Omega \Gamma_1$ with Ω compact hence $H_1 \mathbf{v} = \Omega \Gamma_1 \mathbf{v}$ is closed in $V_{\mathbb{R}}$ proving lemma 3.3.

Lemma 3.3

There exists an $(M_2)_{\mathbb{C}} \times (M_1)_{\mathbb{C}}$ invariant element f of $E[G]$ (the ring of regular functions on M) so that

$$\begin{aligned}f(x) &= 0 & x \in (H_2)_{\mathbb{C}}(H_1)_{\mathbb{C}}, \\ f(x) &= 1 & (Hx \in_2)_{\mathbb{C}} \gamma_j (H_1)_{\mathbb{C}} \quad \gamma_j \in I',\end{aligned}$$

where I' is the set of elements $\gamma_j \in I$ so that

$$(H_2)_{\mathbb{C}} \gamma_j (H_1)_{\mathbb{C}} \neq (H_2)_{\mathbb{C}} (H_1)_{\mathbb{C}}.$$

Proof: It follows from Borel [3], 7.6, that such an f exists with $f \in \mathbb{C}[G]$. However, the regular functions in $E[G]$ invariant under $(M_2)_E \times (M_1)_E$ span the regular functions in $\mathbb{C}[G]$ invariant under $(H_2)_{\mathbb{C}} \times (H_1)_{\mathbb{C}}$. Thus for each $V_j (j \in I')$ we may find f_j so that

$$(i) f_j \in E[G],$$

$$(ii) f_j(x) = 0 \quad x \in (H_2)_{\mathbb{C}} \gamma (H_1)_{\mathbb{C}} \neq (H_2)_{\mathbb{C}} \gamma_j (H_1)_{\mathbb{C}}, \gamma \in I$$

$$(iii) f_j(x) = 1 \quad x \in (H_2)_{\mathbb{C}} \gamma_j (H_1)_{\mathbb{C}}.$$

Then $f = \sum_{j \in I'} f_j$ satisfies the conclusions of lemma 3.4.

Theorem 3.1 now follows easily. If p is a prime which does not divide the denominators of f then we claim that if $\gamma \in I(p)$ then $\gamma \in (H_2)_{\mathbb{C}} (H_1)_{\mathbb{C}}$. Suppose that $\gamma \in I(p)$ and $\gamma \notin (H_2)_{\mathbb{C}} (H_1)_{\mathbb{C}}$. Since $\gamma \in I$ there exist $\gamma_2 \in \Gamma_2$ and $\gamma_1 \in \Gamma_1$ so that $\gamma_2 \gamma \gamma_1 \in I$, hence $\gamma \in (H_2)_{\mathbb{C}} \gamma_j (H_1)_{\mathbb{C}}$ for some $j \in I'$. But then we have $f(\gamma) = 1$. However, $\gamma \equiv 1 \pmod{p}$ so we must have $f(\gamma) \equiv 0 \pmod{p}$ which contradicts $f(\gamma) = 1$. Thus theorem 3.1 is proved.

We will now rename $\Gamma(p)$ as Γ and $I(p)$ as I for a chosen p so that the conclusions of theorem 3.1 are satisfied. Thus $\gamma \in I$ implies $\gamma \in (H_2)_{\mathbb{C}} (H_1)_{\mathbb{C}}$.

If $\gamma \in I$ and $\gamma \in H_2^0 H_1^0$ then the intersection number corresponding to γ has multiplicity $+1$ according to lemma 3.1. Thus we must study the problem. Given $\gamma \in \Gamma$, satisfying $\gamma \in (H_2)_{\mathbb{C}} (H_1)_{\mathbb{C}}$ can we obtain $\gamma \in H_2 H_1$. This problem is a problem in Galois cohomology which we now treat.

Theorem 3.1 allows us to associate a Galois cohomology class $a(\gamma) \in H^1(E, M_1 \cap M_2)$ to every intersection $\gamma \in I$. Indeed, by theorem 3.1 we may write $\gamma = g_2 g_1$ with $g_1 \in M_1(\mathbb{C})$ and $g_2 \in M_2(\mathbb{C})$. Now if $\sigma \in \text{Gal}(\mathbb{C}, E)$ then $\sigma(\gamma) = \gamma$ and consequently

$$g_1 \sigma(g_1)^{-1} = g_2^{-1} \sigma(g_2).$$

We define the value of the Galois cocycle $a(\gamma)$ at σ , to be denoted a_σ , to be this common value. We then have the following standard lemma.

Lemma 3.4 A

We may write $\gamma = \mu_2 \mu_1$ with $\mu_1 \in M_1(E)$ and $\mu_2 \in M_2(E)$ if and only if the class of the cocycle $a(\gamma)$ in $H^1(E, M_1 \cap M_2)$ is trivial.

Since we have identified E with a subfield of $\mathbb{R}$ we have a map

$$H^1(E, M_1 \cap M_2) \rightarrow H^1(\mathbb{R}, M_1 \cap M_2).$$

We let $a \in H^1(\mathbb{R}, M_1 \cap M_2)$ be the image of a . Then we have the following standard lemma.

Lemma 3.4 B

We may write $\gamma = h_2 h_1$ with $h_1 \in M_1(\mathbb{R})$ and $h_2 \in M_2(\mathbb{R})$ if and only if the class of a in $H^1(\mathbb{R}, M_1 \cap M_2)$ is trivial.

Remark : In this case the intersection corresponding to γ is isolated. We note that a is in the kernel of the mapping

$$H^1(\mathbf{R}, M_1 \cap M_2) \rightarrow H^1(\mathbf{R}, M_1) \times H^1(\mathbf{R}, M_2),$$

induced by the inclusions.

Proposition 3.1

Let Φ denote the set of finite primes of E and for $p \in \Phi$ let E_p denote the completion of E at p . Then we may assume a is in the kernel of the map

$$H^1(E, M_1 \cap M_2) \rightarrow \prod_{p \in \Phi} H^1(E_p, M_1 \cap M_2).$$

Moreover, if all elements of Γ satisfy the congruence $\gamma \equiv 1 \pmod{a}$ for some ideal a then for every finite prime p we may write $\gamma = \mu_2 \mu_1$ with μ_2 and μ_1 in the set of p -adic integral points of M_2 and M_1 respectively and satisfying $\mu_2 \equiv \mu_1 \equiv 1 \pmod{a}$.

Proof : Since the second part of the statement of the theorem implies the first we prove this part. Note first that if we may solve $\gamma \equiv \mu_2^{(n)} \mu_1^{(n)} \pmod{p^n}$ with $\mu_1^{(n)}$ and $\mu_2^{(n)}$ p -adic integral matrices congruent to 1 modulo a^1 (depending on n) for every n , then, since the set of p -adic integral matrices is compact we may find (by taking a limit) p -adic integral matrices μ_1, μ_2 congruent to 1 modulo a so that $\gamma = \mu_2 \mu_1$. Hence if the second statement does not hold for some finite prime p then there exists n so that $\gamma \not\equiv \mu_2 \mu_1 \pmod{p^n}$ for any p -adic integral matrices μ_2, μ_1 congruent to 1 modulo a . Now consider the congruence subgroup $\Gamma(p^n)$ of Γ . We claim that there is no $\gamma' \in \Gamma(p^n)$ which lies over γ . Indeed suppose such a γ' existed. Then $\gamma' = \gamma_2 \gamma \gamma_1$ with $\gamma_2 \in \Gamma_2, \gamma_1 \in \Gamma_1$ and $\gamma_2 \equiv 1 \pmod{a}, \gamma_1 \equiv 1 \pmod{a}$ and $\gamma' \equiv 1 \pmod{p^n}$. Consequently $\gamma \equiv \gamma_2^{-1} \gamma_1^{-1} \pmod{p^n}$ but taking $\mu_2 = \gamma_2^{-1}$ and $\mu_1 = \gamma_1^{-1}$ we obtain a contradiction. In this way we eliminate all intersections not satisfying the conclusion of proposition 3.1.

Theorem 3.2

Suppose the following two hypotheses are satisfied:

$$(1) H^1(\mathbf{R}, M_1 \cap M_2) \rightarrow H^1(\mathbf{R}, H_1) \times H^1(\mathbf{R}, H_2),$$

is injective—under this assumption we obtain that a is trivial.

(2) The Hasse principle holds for $M_1 \cap M_2$; that is, the map

$$H^1(E, M_1 \cap M_2) \rightarrow \prod_v H^1(E_v, M_1 \cap M_2)$$

induced by the various completions is injective (the inverse image of the base-point consists of a single point) then we may write each γ in I as a product $\gamma = m_2 m_1$ with $m_1 \in M_1(E)$ and $m_2 \in M_2(E)$.

Proof : Since a is trivial we may write $\gamma = h_2 h_1$ with $h_2 \in H_2$ and $h_1 \in H_1$. But if $M_1^{\tau_i}$ is a conjugate of M_1 then $M_1^{\tau_i}(\mathbf{R})$ is compact and consequently the mapping $H^1(\mathbf{R}, M_1^{\tau_i} \cap M_2^{\tau_i}) \rightarrow H^1(\mathbf{R}, M_1^{\tau_i})$ is injective, see Serre [12], page III-36,

theorem 6. Consequently a is trivial at all the infinite completions of E . But by proposition 3.1 we know a is trivial at all the finite completions of E ; hence, by the Hasse principle a is trivial and we may write $\gamma = m_2 m_1$ with $m_2 \in M_2(E)$ and $m_1 \in M_1(E)$.

Remark: In the special case that both H_1 and H_2 are connected then (1) is sufficient to guarantee that all intersections of Y_1 and Y_2 are isolated of multiplicity $+1$.

4. Applications

We will apply our results to three cases. In all three cases we begin with a real quadratic field $E = \mathbb{Q}(\sqrt{m})$ where m is a rational prime.

In the first case we let D be a division quaternion algebra over E so that $D \otimes_E \mathbb{R}$ is the Hamilton quaternions H . We define a hermitian form H on D^n by the formula

$$H(q_1, q_2, \dots, q_n) = q_1 \bar{q}_1 + \dots + q_p \bar{q}_p - \sqrt{m}(q_{p+1} \bar{q}_{p+1} + \dots + q_n \bar{q}_n),$$

where $\bar{q}$ denotes the conjugate of the quaternion q . We assume $p \geq q$ and choose an integer k satisfying $0 < k < p$. We let M be the special unitary group of H ; we let M_1 be the subgroup of M fixing the first k coordinates and we let M_2 be subgroup of M fixing the next $p - k$ coordinates. Then M, M_1 and M_2 are algebraic groups defined over E and

$$M(\mathbb{R}) = Sp(p, q),$$

$$M_1(\mathbb{R}) = Sp(p - k, q),$$

$$M_2(\mathbb{R}) = Sp(k, q).$$

Since H is anisotropic with a positive definite conjugate it follows by a standard argument, see for example Millson [10], page 239, that $M(O_E)$ is a uniform discrete subgroup of $Sp(p, q)$.

In the second case we let F be a quadratic extension of E so that $F \otimes_E \mathbb{R} = \mathbb{C}$. We define a hermitian form H on F^n by the formula

$$H(z_1, z_2, \dots, z_n) = z_1 \bar{z}_1 + \dots + z_p \bar{z}_p - \sqrt{m}(z_{p+1} \bar{z}_{p+1} + \dots + z_n \bar{z}_n),$$

where $\bar{z}$ denotes the conjugate of z by the non-trivial element of the Galois group of F over E . The definitions of M, M_1 and M_2 are now as in the previous case. We now have

$$M(\mathbb{R}) = SU(p, q),$$

$$M_1(\mathbb{R}) = SU(p - k, q),$$

$$M_2(\mathbb{R}) = SU(k, q).$$

Again we find that $M(O_E)$ is a uniform discrete subgroup of $SU(p, q)$.

In the third case we let M be the orthogonal group of the quadratic form f over E defined by

$$f(x_1, x_2, \dots, x_n) = x_1^2 + \dots + x_p^2 - \sqrt{m}(x_{p+1}^2 + \dots + x_n^2).$$

The groups M_1 and M_2 are defined to be the subgroups of M_1 and M_2 leaving fixed the first k coordinates and next $p - k$ coordinates respectively. In this case we have

$$M(\mathbf{R}) = SO(p, q),$$

$$M_1(\mathbf{R}) = SO(p - k, q),$$

$$M_2(\mathbf{R}) = SO(k, q).$$

We find that $M(O_E)$ is a uniform discrete subgroup of $SO(p, q)$.

The first two cases can be settled by the remark following theorem 3.3. The mapping $H_1 \cap H_2 \rightarrow H_1$ is the inclusion of one unitary group into a larger one, consequently by Witt's theorem the mapping $H^1(\mathbf{R}, M_1 \cap M_2) \rightarrow H^1(\mathbf{R}, M_1)$ is injective. Since H_1 and H_2 are connected we obtain the following applications (note that our construction may be repeated interchanging the roles of p and q).

Theorem 4.1

There exists a uniform discrete arithmetic subgroup $\bar{\Gamma} \subset Sp(p, q)$ so that $H^k(\bar{\Gamma}, \mathbf{R})$ contains a cohomology class which is not the restriction of a continuous class from $Sp(p, q)$ for any integer k strictly between 0 and $4pq$ and divisible by either $4p$ or $4q$.

Theorem 4.1 (bis)

There exists a uniform discrete arithmetic subgroup $\Gamma \subset SU(p, q)$ so that $H^k(\Gamma, \mathbf{R})$ contains a cohomology class which is not the restriction of a continuous class of $SU(p, q)$ for any integer k strictly between 0 and $2pq$ and divisible by either $2p$ or $2q$.

Remark : In the case $q = 1$ the symmetric space we are looking at is the unit ball in $\mathbf{C}^n$, the cycles are quotients of sub-balls dual to linear subspaces of $\mathbf{CP}^n$. This justifies the statement (3) of the introduction.

The remaining case of $M(\mathbf{R}) = SO(p, q)$ is considerably more difficult. The difficulty is that we cannot find a simple algebraic subgroup N acting transitively on X_1 so that $N(\mathbf{R})$ is connected—note that the group $SO(p, q)$ has two components. Our first problem is to ensure that the manifolds Y , Y_1 and Y_2 are orientable. To do this we must ensure that Γ_1 and Γ_2 are contained in the connected components of the identity of G_1 and G_2 respectively. We will show in fact that there exists an ideal $\mathfrak{a}$ in E so that $\gamma \equiv 1 \pmod{\mathfrak{a}}$ implies that $\theta(\gamma) = 1$ where θ is the spinorial norm which we now define as it will play a major role in what follows.

Definition : Given a quadratic form f on a finite dimensional vector space over a field K we define a homomorphism $\theta : O(f) \rightarrow [K^*/(K^*)^2]$ as follows. Given $u \in O(f)$ write u as a product of reflections $u = r_{x_1} r_{x_2} \dots r_{x_n}$ where r_{x_i} is reflection in the vector x_i . This can always be done—see O'Meara [9], 43.3. Then

$$\theta(u) = f(x_1)f(x_2)\dots f(x_n) \pmod{(K^*)^2}.$$

We define $O'(f)$ to be the subgroup of $SO(f)$ consisting of elements of spinorial norm 1. We remark that if $SO(p, q)$ is the group of orientation preserving

isometries of a quadratic form over $\mathbb{R}$ of signature (p, q) then the connected component of the identity of $SO(p, q)$ is just $O'(p, q)$. We note $u \in O'(p, q)$ if and only if we may write $u = r_{x_1} r_{x_2} \dots r_{x_n}$ with $f(x_1) f(x_2) \dots f(x_n) > 0$. If $q = 1$ we see that $u \in O'(p, 1)$ if and only if u maps the half-space $x_q > 0$ into itself.

We will use the following notation. If γ is a p -adic integral matrix and b is an ideal in O then $\gamma \equiv 1 \pmod{b}$ means $\gamma \equiv 1 \pmod{p^{v(b)}}$ where v is the p -adic valuation.

Proposition 4.1

Let Q be a quadratic form over a number field E . Let $S = \{p_1, p_2, \dots, p_m\}$ be a finite set of primes of E and let $G(O_S)$ be the group of S integral matrices over E that preserve Q . Then there exist infinitely many relatively prime ideals O in O_S with the property that every element of $G(O_S)$ congruent to 1 modulo a has spinorial norm 1 in E .

Proof: We denote $G(O_S)$ by Γ .

We first consider a prime p in E so that p is not a divisor of 2 and so that p does not divide the discriminant of Q . In this case if E_p denotes the completion of E at p we obtain a form Q_p defined over E_p with orthogonal group $O(Q_p)$. We have inclusions

$$\begin{array}{ccc} \Gamma_1(\mathcal{G}) & \xrightarrow{\quad} & O(Q_p) \\ \downarrow & \nearrow & \\ \Gamma_1 & & \end{array}$$

We denote the p -adic spinorial norm by θ_p . The following lemma was kindly provided for us by T. Tamagawa.

Lemma 5.3

If $\gamma \in \Gamma_1(p)$ then in $O(Q_p)$ we may write $\gamma = r_{x_1} o_{x_2} \dots o_{x_m}$ with $Q(x_j)$ a p -adic unit, $1 \leq j \leq m$.

Proof: The proof is by induction. We establish the case $n = 2$ and the induction step with the same argument. Let $\gamma \in \Gamma_1(p)$ and suppose $x \in E_p^n$ with $Q(x)$ a p -adic unit. Let $y = \gamma x$. Let $\rho = r_{x+y} o_{x_2}$, then $\rho x = y$ and hence $\rho^{-1} \gamma(x) = x$. But $Q(x+y) \equiv 4Q(x) \pmod{p}$ and accordingly is a unit.

Let U_p denote the p -adic units. We have established that there exists a homomorphism $\tilde{\theta}_p$ satisfying

$$\begin{array}{ccc} \Gamma(\mathcal{G}) & \xrightarrow{\tilde{\theta}_{\mathcal{G}}} & \frac{U_{\mathcal{G}}}{U_{\mathcal{G}}^2} \\ & \searrow \theta_{\mathcal{G}} & \\ & & \frac{E^x}{(E^x_{\mathcal{G}})^2} \end{array}$$

But since $\gamma \equiv 1 \pmod p$ we have $\theta(\gamma)$ is a square modulo p hence a square in U_p by Hensel's lemma. We obtain then

Lemma 5.4

The restriction of p -adic spinorial norm to $\Gamma_1(p)$ is trivial.

We are now ready to finish the proof of proposition 4.1. We know

$$\theta(\Gamma_1) \subseteq \frac{E^*}{(E^*)^2}$$

is a finitely-generated abelian group of exponent 2. Let $u_1, u_2, \dots, u_r$ be a set of generators. Then $u \in \theta(\Gamma_1)$ has the form

$$u = u_1^{e_1} u_2^{e_2} \dots u_r^{e_r}$$

with e_j either 0 or 1 for $1 \leq j \leq r$. Since u_j is not a square in E there exist infinitely many p -adic fields Ep so that u_j is not a square in Ep . Let Ep_j be such and satisfying further that u_j is not a square mod Ep_k for $k \neq j$, one of these for each j between 1 and r . Then by construction we have an injection i

$$\begin{array}{ccc} \theta(\Gamma_1) & \xrightarrow{i} & \frac{E_{\mathcal{G}_1}^*}{(E_{\mathcal{G}_1}^*)^2} \times \frac{E_{\mathcal{G}_2}^*}{(E_{\mathcal{G}_2}^*)^2} \times \dots \times \frac{E_{\mathcal{G}_r}^*}{(E_{\mathcal{G}_r}^*)^2} \\ & \searrow & \uparrow \\ & E^* & \\ & \frac{(E^*)^2}{} & \end{array}$$

Indeed $u = u_1^{e_1} \dots u_r^{e_r} \equiv x^2 \pmod{p_1 \dots p_r} \Leftrightarrow u_j^{e_j} \equiv x^2 \pmod{p_j}$ for all j .

Lemma 5.5

$$\theta(\Gamma_1(p_1, p_2, \dots, p_r)) = \frac{(E^*)^2}{(E^*)^3}.$$

Proof: Suppose that $u \in \theta(\Gamma_1(p_1, p_2, \dots, p_r))$. To show u is 1 it is sufficient to show $i(u)$ is 1. But we have a commutative diagram

$$\begin{array}{ccc} \Gamma_1(\mathcal{G}_1, \dots, \mathcal{G}_r) & & \prod_{j=1}^r \theta_{\mathcal{G}_j} \\ \downarrow & \searrow & \uparrow \\ \frac{E^*}{(E^*)^2} & \xrightarrow{i} & \prod_{j=1}^r \frac{E_{\mathcal{G}_j}^*}{(E_{\mathcal{G}_j}^*)^2} \end{array}$$

and by lemma 5.2 each $\theta_{p_j} | \Gamma(p_1, \dots, p_r)$ is trivial.

Corollary 1

There exist infinitely many relatively prime ideals $\mathfrak{a}$ in $\mathcal{O}$ so that $\gamma \equiv 1 \pmod{\mathfrak{a}}$ implies $\theta(\gamma) = 1$.

We will choose an ideal $\mathfrak{a}$ satisfying the conditions: $\gamma \equiv 1 \pmod{\mathfrak{a}}$ implies γ is not of finite order, $\theta(\gamma) = 1$ and so that theorem 3.1 is satisfied. $\mathfrak{a}$ will henceforth denote such an ideal and $\Gamma = \{\gamma \in \Phi : \gamma \equiv 1 \pmod{\mathfrak{a}}\}$.

Corollary 2

Y , Y_1 , and Y_2 are orientable.

We are now ready to apply our general theory. $M_1 \cap M_2$ is an orthogonal group and consequently satisfies the Hasse principle. The mapping $H_1 \cap H_2 \rightarrow H_1$ is the inclusion of an orthogonal group into a bigger orthogonal group, consequently, by Witt's theorem the mapping $H^1(\mathbf{R}, M_1 \cap M_2) \rightarrow H^1(\mathbf{R}, M_1)$ is injective. We apply theorem 3.2 and find that by passing to a deeper congruence subgroup which we will continue to denote by Γ we obtain, if $\gamma \in I(\Gamma)$ then $\gamma = m_2 m_1$ with $m_1 \in M_1(E)$ and $m_2 \in M_2(E)$.

We now prove the first claim of the introduction. For this we use the groups discussed under case 3 at the beginning of this chapter with $q = 1$ and k any integer strictly between 0 and p . We will verify that $Y_1 \cap Y_2$ is a single point. Indeed let $\gamma, m_2 m_1$ be as above. We will show that the equation $\gamma = m_2 m_1$ determines all the entries of m_2 . Of course, $m_2 e_j = e_j$ for $k < j < p + 1$. Moreover

$$\gamma e_j = m_2 m_1 e_j = m_2 e_j \text{ for } 1 \leq j \leq k.$$

Thus it is immediate that m_2 is determined by γ on all basis vectors with the possible exception of the last vector e_{p+1} . We suppose that $m_2 e_{p+1} = u$. Then u is in the orthogonal complement of the vectors $\gamma e_1, \gamma e_2, \dots, \gamma e_k$ in the space spanned by $e_1, e_2, \dots, e_k, e_{p+1}$. Consequently u is determined up to a scalar multiple which is either ± 1 because

$$f(u) = f(m_2 e_{p+1}) = f(e_{p+1}) = -\sqrt{m}.$$

In other words u is determined up to multiplication by ± 1 by the above k linear equations and the quadratic equation $f(u) = -\sqrt{m}$. We denote this set of $k + 1$ equations by $*$. Now given any prime p we have the decomposition $\gamma = v_p \mu_p$ of proposition 3.1. But an argument identical to the above shows that $v_p e_{p+1}$ is determined up to sign by the same system $*$. Consequently $u = \pm v_p e_{p+1}$ and consequently u is integral. We have found then that m_2 is in fact an integral matrix in $M_2(O)$. We reduce m_2 modulo $\mathfrak{a}$ and find that since $\gamma e_j \equiv e_j \pmod{\mathfrak{a}}$ $1 \leq j \leq p$ we must have $m_2 e_{p+1} \equiv \pm e_{p+1} \pmod{\mathfrak{a}}$ but the determinant of m_2 is 1; hence, $m_2 \equiv 1 \pmod{\mathfrak{a}}$. This result combined with results in Borel-Wallach [5] proves the first claim of the introduction.

We now prove the second claim of the introduction. We use the groups M_1, M_2, M discussed under case 3 at the beginning of this chapter with q arbitrary and $k = 1$. From our general theory, given $\gamma \in I$ we may write $\gamma = m_2 m_1$ with $m_2 \in M_2(E)$ and $m_1 \in M_1(E)$. Since $\theta(\gamma) = 1$ we have $\theta(m_2) = \theta(m_1)$. Thus either $\theta(m_2)$ and $\theta(m_1)$ are totally positive or we may rewrite $\gamma = m'_2 \tau_2 \tau_1 m'_1$ with $\theta(m'_2)$ and $\theta(m'_1)$ totally positive, τ_2 the diagonal matrix with first entry -1 , $p + q$ th entry -1 and the rest $+1$ and τ_1 the diagonal matrix with second diagonal entry -1 , $p + q$ th entry -1 and the rest $+1$. We now examine under what conditions γ corresponds to a positive intersection. In case $\theta(m_1)$ and $\theta(m_2)$ are both positive then $\gamma \in H_2^0 H_1^0$ and of course $\epsilon(\gamma) = 1$. To compute $\epsilon(\gamma)$ in the second case we have only to compute

determinant of the action of $\tau_2\tau_1$ on Φ_1 . But it is easily seen that $\tau_2\tau_1 = -1$ on Φ_1 and consequently has determinant $(-1)^q$. This completes the proof of the second claim in the introduction.

In connection with the third claim of the introduction we note that for $n \geq 2$, for any choice of k , the cycles Y_1 and Y_2 are projective varieties. Since (using Witt's theorem) we have shown that all intersections are isolated, we have shown that all intersections of algebraic cycles in all even dimensions which correspond to non-trivial representations are isolated.

References

- Birkes D 1971 Orbits of linear algebraic groups. *Ann. Math.* (2) 93 459-475
- Borel A 1976 Cohomologie de sous-groupes discrets et représentations de groupes semi-simples. *Bull. Soc. Math. France* 32-33 73-112
- Borel A 1969 *Introduction aux Groupes Arithmétiques* (Hermann)
- Borel A 1963 Compact Clifford-Klein forms of symmetric spaces. *Topology* 2 111-122
- Borel A and Wallach N Seminar on the cohomology of discrete subgroups of semi-simple groups. *Ann. Math. Studies*, Princeton University Press (to appear)
- Casselman W and Schmid W to appear
- Helgason S *Differential geometry and symmetric spaces* (New York: Academic Press)
- Kneser M 1966 Strong approximation. *Proc. Symp. Pure Math.* 9 187-196
- O'Meara O T 1971 *Introduction to quadratic forms* (New York: Springer Verlag)
- Millson J J 1976 On the first Betti number of a constant negatively curved manifold. *Ann. Math.* 104 235-247
- Mostow G D 1973 Strong rigidity of locally symmetric spaces. *Ann. Math. Studies* (Princeton, New Jersey)
- Serre J P 1973 *Cohomologie Galoisienne* (Springer-Verlag)
- Zuckerman G Continuous cohomology and unitary representations of real reductive groups (to appear)
- Tits J 1966 Classification of algebraic semisimple groups. *Proc. Symp. Pure Math.* 9 33-62

University
Haven,
Connecticut 06520, USA

Institute of Fundamental Research,
Hay 400 005

the determinant of the action of $\tau_2\tau_1$ on Φ_1 . But it is easily seen that $\tau_2\tau_1 = -1$ on Φ_1 and consequently has determinant $(-1)^a$. This completes the proof of the second claim in the introduction.

In connection with the third claim of the introduction we note that for $SO(n, 2)$, for any choice of k , the cycles Y_1 and Y_2 are projective varieties. Since (by using Witt's theorem) we have shown that all intersections are isolated, we obtain algebraic cycles in all even dimensions which correspond to non-trivial automorphic representations.

References

- [1] Birkes D 1971 Orbits of linear algebraic groups. *Ann. Math.* (2) **93** 459–475
- [2] Borel A 1976 Cohomologie de sous-groupes discrets et représentations de groupes semi-simples. *Bull. Soc. Math. France* **32–33** 73–112
- [3] Borel A 1969 *Introduction aux Groupes Arithmétiques* (Hermann)
- [4] Borel A 1963 Compact Clifford-Klein forms of symmetric spaces. *Topology* **2** 111–122
- [5] Borel A and Wallach N Seminar on the cohomology of discrete subgroups of semi-simple groups. *Ann. Math. Studies*, Princeton University Press (to appear)
- [6] Casselman W and Schmid W to appear
- [7] Helgason S *Differential geometry and symmetric spaces* (New York: Academic Press)
- [8] Kneser M 1966 Strong approximation. *Proc. Symp. Pure Math.* **9** 187–196
- [9] O'Meara O T 1971 *Introduction to quadratic forms* (New York: Springer Verlag)
- [10] Millson J J 1976 On the first Betti number of a constant negatively curved manifold. *Ann. Math.* **104** 235–247
- [11] Mostow G D 1973 Strong rigidity of locally symmetric spaces. *Ann. Math. Studies* (Princeton, New Jersey)
- [12] Serre J P 1973 *Cohomologie Galoisienne* (Springer-Verlag)
- [13] Zuckerman G Continuous cohomology and unitary representations of real reductive groups (to appear)
- [14] Tits J 1966 Classification of algebraic semisimple groups. *Proc. Symp. Pure Math.* **9** 33–62

Yale University
 New Haven,
 Connecticut 06520, USA
 and
 Tata Institute of Fundamental Research,
 Bombay 400 005

Polarisations on an abelian variety

By

M S NARASIMHAN and M V NORI

1. Statement of the theorem and preliminaries

We shall prove the following.

Theorem 1.1. Let X be an abelian variety over an algebraically closed field k and let $\text{Aut}(X)$ denote the automorphism group of X . If, for any integer $d \neq 0$, P_d denotes the subset of the Neron-Severi group of X consisting of elements of degree d , then there are only finitely many orbits under the natural action of $\text{Aut}(X)$ on P_d .

For the definition of degree, see § 1.6.

In particular, if P denotes the set of principal polarisations on X , there are only finitely many orbits under the natural action of $\text{Aut}(X)$ on P . This result was conjectured in the case $k = \mathbb{C}$, by Martens ([3], p. 47).

On applying Torelli's theorem we have

Corollary 1.2. There are, upto isomorphism, only finitely many smooth irreducible curves over k , having a given abelian variety as the Jacobian.

In the proof of theorem 1.1, we will use

Theorem 1.3. (Borel and Harish-Chandra) Let G be a reductive algebraic group defined over $\mathbb{Q}$ and let Γ be an arithmetic subgroup of G [2]. Let $\rho : G \rightarrow GL(V)$ be a rational representation defined over $\mathbb{Q}$ and L a lattice of V invariant under Γ . If O is a closed orbit of G , then $O \cap V$ consists of a finite number of orbits of Γ .

For a proof see ([1], theorem in § 9.11) or ([2], theorem 6.9).

1.4. Aut(X) as an arithmetic group. Let B denote the ring of endomorphisms of X and let $B_K = B \otimes K$ for any field K containing $\mathbb{Q}$. Then $B_{\mathbb{Q}}$ is a finite dimensional semi-simple algebra which contains B as a lattice ([4], theorem 3 on p. 176; also p. 178). Moreover there is a homogeneous polynomial function, with rational coefficients, ϕ defined over B such that $\phi(ab) = \phi(a)\phi(b)$ for $a, b \in B$ and $\phi(a) = \deg a$, for $a \in B$, where $\deg a$ denotes the degree of the endomorphism a ([4], theorem 2, p. 174; also p. 63).

Lemma 1.5. Let $\alpha \in B_K$, with $\phi(\alpha) \neq 0$. Then α is a unit of B_K .

Proof. For $\alpha \in B$, $\phi(\alpha) = \deg \alpha = n \neq 0$, there is an element β of B such that $\alpha\beta = \beta\alpha = n_{\mathbf{x}}$ ([4], p. 169) proving that α is a unit of $B_{\mathbb{Q}}$.

Because ϕ is homogeneous, this proves the lemma for $K = \mathbb{Q}$. The general case follows from this. In fact, let $B = \prod S_i$, where S_i are simple algebras over

Q. It then follows that $\phi = \prod N_i^{k_i}$ with $k_i > 0$ where N_i is the reduced norm in S_i . Since the determinant of the regular representation of $S_i \otimes_{\mathbf{Q}} K$ is a positive power of $N_{i,K}$ (the extension of N_i to $S_i \otimes_{\mathbf{Q}} K$), we see that an element a of $S_i \otimes K$ is a unit if and only if $N_{i,K}^{k_i}(a) \neq 0$. Thus a is a unit if and only if $\phi(a) \neq 0$.

Putting

$$G_K = \{a \in B_K \mid \phi(a) = 1\}$$

for all fields K containing $\mathbf{Q}$, one gets a reductive algebraic group G defined over $\mathbf{Q}$, as $B_{\mathbf{Q}}$ is a semi-simple algebra. Moreover G contains

$$\text{Aut}(X) = \{a \in B \mid a \text{ is unit}\} = \{a \in B \mid \phi(a) = \deg a = 1\}$$

as an arithmetic group.

1.6. Rosati involution and the action of $\text{Aut}(X)$ on the Neron-Severi group

In this section, we construct a representation of G defined over $\mathbf{Q}$, which when restricted to the arithmetic subgroup $\text{Aut}(X)$ becomes the natural representation of $\text{Aut}(X)$ on the Neron-Severi group.

Let $NS(X) = \text{Pic } X / \text{Pic}^{\circ} X$

denote the Neron-Severi group of X . Let $n \in NS(X)$ and L a line bundle representing n . The homomorphism

$$\phi_L: X \rightarrow \text{Pic}^{\circ} X = \hat{X} \text{ (the dual abelian variety)}$$

defined by $\phi_L(x) = T_x^*(L) \otimes L^{-1}$ depends only on n and will be denoted by ϕ_n . We define

$$\deg n = \begin{cases} \deg \phi_n & \text{if } \phi_n \text{ is an isogeny} \\ 0 & \text{otherwise.} \end{cases}$$

Fix, once for all, an element L_0 in $NS(X)$ corresponding to an ample line bundle. Let $d_0 = \deg L_0$. For any endomorphism f of X , put

$$f' = \phi_{L_0}^{-1} \circ \hat{f} \circ \phi_{L_0} \in B_{\mathbf{Q}},$$

where $\hat{f}: \hat{X} \rightarrow \hat{X}$ in the transpose of f . The map $f \mapsto f'$ extends to an involution $\theta_{\mathbf{Q}}$ of $B_{\mathbf{Q}}$, i.e., $\theta_{\mathbf{Q}}$ is a linear isomorphism of $B_{\mathbf{Q}}$, $\theta_{\mathbf{Q}}^2 = \text{identity}$, $\theta_{\mathbf{Q}}(ab) = \theta_{\mathbf{Q}}(b)\theta_{\mathbf{Q}}(a)$, and $\theta_{\mathbf{Q}}(a) = a'$ for $a \in B$ ($\theta_{\mathbf{Q}}$ is called the Rosati involution associated to L_0). Let $\theta_K = \theta_{\mathbf{Q}} \otimes \text{Id}_K$; then θ_K is an involution of B_K and the $\theta_K: G_K \rightarrow G_K$, $\mathbf{Q} \subset K$, define a morphism $\theta: G \rightarrow G$.

If $n \in NS(X)$, put $\rho(n) = \phi_{L_0}^{-1} \circ \phi_n$.

Then ρ gives an injection of $NS(X)$ into B and $\rho(NS(X)) = F$ is a lattice in

$$\mathbf{Q} \cdot F = S_{\mathbf{Q}} = \{a \in B_{\mathbf{Q}} \mid \theta_{\mathbf{Q}}(a) = a\} \text{ ([4], § 20, p. 190).}$$

Let $S_K = \{b \in B_K \mid \theta_K(b) = b\}$.

Consider the following representation π of G on S defined over $\mathbf{Q}$: for $g \in G_K$ and $s \in S_K$ put $\pi(g^{-1})s = \theta(g)sg$. For $n \in NS(X)$ and $g \in \text{Aut}(X)$, let $g^*(n)$ denote the pull-back of n by g . We then have

$$\theta(g)\rho(n)g = \phi_{L_0}^{-1} \hat{g} \phi_{L_0} \circ \phi_{L_0}^{-1} \phi_n \circ g = \phi_{L_0}^{-1} \hat{g} \phi_n g = \phi_{L_0}^{-1} \phi_{g^*n} = \rho(g^*n).$$

Thus $\pi(\text{Aut } X)$ leaves the lattice $\rho(NS(X))$ in $B_{\mathbf{Q}}$ invariant and the action π of $\text{Aut}(X)$ on $\rho(NS(X))$ coincides with the natural action of $\text{Aut } X$ on $NS(X)$.

2. Algebras with involution

In this section, we study algebras with involution over an algebraically closed field K of characteristic zero. All the involutions encountered will be denoted by θ .

Let V, W be vector spaces of dimension n over K and $(,): V \times W \rightarrow k$ be a non-degenerate bilinear pairing. If $A: V \rightarrow V$ is an endomorphism, denote by A' the endomorphism of W defined by $(AX, Y) = (X, A'Y)$ for $X \in V, Y \in W$. Then $A \mapsto A'$ gives an isomorphism of $\text{End } V$ onto $\text{End } W$. Put $C_n = \text{End } V \times \text{End } W$ and

$$\theta(A, B) = (B', A').$$

Now if $V = W$ and $A' = A$ for all A in $\text{End } V$, then the bilinear form $(,)$ is either symmetric or skew-symmetric. The algebra $\text{End } V$ with the involution $\theta(A) = A'$ will be denoted by A_n in the first case and by B_n in the second case.

Lemma 2.1. Let $\alpha \in A_n, B_n$ or C_n be a unit such that $\theta(\alpha) = \alpha$; then we may write $\alpha = \beta\theta(j)j$ where β is in the centre and j satisfies:

- (i) if $\alpha \in A_n$ or B_n , $\det(j) = 1$,
- (ii) if $\alpha \in C_n$, $j = (j_1, j_2)$, then $\det j_1 = \det j_2 = 1$.

Proof: If $\alpha \in C_n$, α is of the form (A, A') ; put $A = \beta B$ with $\beta \in K^*$, $\det B = 1$, and let $j = (B, 1)$. Then $\alpha = \beta^2 \theta(j)j$.

If $\alpha \in A_n$ (resp. B_n), the form $(\alpha X, Y)$ is a non-degenerate bilinear symmetric (resp. antisymmetric) form on V , and so is (X, Y) .

Hence there exists $g \in \text{Aut } V$ such that $(\alpha X, Y) = (gX, gY)$ for all $X, Y \in V$. Again, putting $g = \beta j$ with $\beta \in K^*$ and $\det j = 1$, we have $\alpha = \beta^2 \theta(j)j$.

Lemma 2.2. Let B be a semi-simple algebra with involution θ over K . Let α be a unit of B with $\theta(\alpha) = \alpha$. Then we may write $\alpha = \beta\theta(j)j$, where β is in the centre and $N(j) = 1$ for any homogeneous polynomial function N on B satisfying $N(ab) = N(a)N(b)$ for all $a, b \in B$ and $N(1) = 1$.

Proof: It is well known that any such algebra with involution is a product of algebras with involution of the type A_n, B_n or C_n ([5], p. 595]. Write B as such a product. Then N is a product of powers of the determinant functions in the various factors. The lemma now follows from lemma 2.1.

3. Proof of theorem 1.1

In view of theorem 1.3 and the observations in § 1.6, the theorem will be a consequence of

Lemma 3.1. For any algebraically closed field K of characteristic zero, there are only finitely many G_K -orbits in the closed set

$$S_d = \{a \in S_K \mid \phi(a) = d/d_0\},$$

where $S_K = S_{\mathbb{Q}} \otimes_{\mathbb{Q}} K$

and $S_{\mathbb{Q}} = \{a \in B_{\mathbb{Q}} \mid \theta_{\mathbb{Q}}(a) = a\}$.

Moreover, each of these orbits is closed.

Proof: Let Z' be the centre of B_K and $Z = Z' \cap S_K$, $R = G_K \cap Z$. If $a \in S_K$ with $\phi(a) = d/d_0$, ($d \neq 0$), a is a unit by lemma 1.5. Using lemma 2.2, we may write $a = \theta(j)\beta j$ with $\beta \in Z$ and $\phi(j) = 1$.

So, in order to show that there are finitely many G_K -orbits in S_d , it suffices to show that there exist only finitely many R -orbits in $S_d \cap Z$. If $a \in S_d \cap Z$, a is a unit by lemma 1.5 and so $S_d \cap Z = aR$, from which it follows that the R -orbits in $S_d \cap Z$ can be put in bijective correspondence with R/R^2 . Now R is the kernel of $\phi \mid Z^*$ so that R is the product of a torus and a finite abelian group, proving the first part of the lemma.

To prove that the orbits are closed it is enough to show that they are equidimensional or, what is the same, that the isotropy groups in question have the same dimension. For this, it is enough, by the preceding considerations, to look at the isotropy groups of points in $S_d \cap Z$. But here any isotropy group clearly consists of elements $g \in G_K$ such that $\theta(g)g = \text{identity}$.

References

- [1] Borel A 1969 *Introduction aux groupes arithmétiques* (Paris : Hermann)
- [2] Borel A and Harish-Chandra 1962 Arithmetic subgroups of algebraic groups; *Ann. Math.* 75 485-535
- [3] Martens H Riemann matrices with many polarisations, *Complex analysis and its applications*, Vol. III (ICTP Vienna: Summer School, Int. Atomic Energy Agency), pp. 35-48
- [4] Mumford D 1969 *Abelian varieties*, *T.I.F.R. studies in mathematics* (Oxford University Press)
- [5] Weil A 1960 Algebras with involution and classical groups; *J. Indian Math. Soc.* 24 591

Poisson formulae of Hecke type

By

S RAGHAVAN and S S RANGACHARI

1. Introduction

In his lectures on "Forms of higher degree", Igusa [10] gave a simple proof of a Poisson formula due to Yamazaki [19] associated with coefficients of Dirichlet series with functional equations involving a single Γ -factor, by introducing an operator of order 2 in a function space, via the Mellin transform. Our object here is to show that this Poisson formula can be generalised a little, so as to bring within its ambit, functional equations of Dirichlet series involving multiple Γ -factors (and, in particular, those associated with Mellin transforms of non-analytic automorphic functions). We also indicate an adelic interpretation of certain Poisson formulae above just to highlight the fact that such a formula in adelic form constitutes an important step in the proof of theorem 5 in § 3 of [7] on a global representation of $(GL_2)_A$.

2. Mellin transforms

Let $\mu_{0,1}, \dots, \mu_{0,a}, \mu_{1,1}, \dots, \mu_{1,a}, \mu_{2,1}, \dots, \mu_{2,a}$

be a sequence of a -tuples of complex numbers such that $\text{Re}(\mu_{k,j}) = \lambda_k$ for every j with $1 \leq j \leq a$ and for every $k \geq 0$ and further let $0 \leq \lambda_0 < \lambda_1 < \lambda_2 \dots \rightarrow \infty$. For every $k \geq 0$, let $m_k \geq 1$ be a fixed natural number.

Following Igusa [10], we define corresponding function spaces $\mathcal{F}$ and $\mathcal{X}$ by the conditions:

(A) $\mathcal{F}$ consists of all complex valued C^∞ functions F on the space $\mathbb{R}_+^\times$ of positive real numbers behaving like Schwartz functions at infinity such that with well-determined complex constants $a_{k,j,m}$ the function F has a termwise differentiable asymptotic expansion

$$F(x) \approx \sum_{0 \leq k < \infty} \sum_{1 \leq j \leq a} \sum_{1 \leq m \leq m_k} a_{k,j,m} x^{\mu_{k,j}} (\log x)^{m-1}, \quad (1)$$

as x tends to 0, and

(B) $\mathcal{Z}$ consists of all complex valued functions Z meromorphic on $\mathbb{C}$ with poles at most at the points $-\mu_{k,j}$ ($k \geq 0$; $1 \leq j \leq a$) and having principal part at $-\mu_{k,j}$ of the form.

$$\sum_{1 \leq m \leq m_k} b_{k,j,m} (s + \mu_{k,j})^{-m}$$

and further such that for every polynomial P in s and any vertical strip $B_{\sigma_1, \sigma_2} = \{s = \sigma + ti \in \mathbb{C}; \sigma_1 \leq \sigma \leq \sigma_2\}$ with arbitrary σ_1, σ_2 , the function $P(s)Z(s)$ is bounded in B_{σ_1, σ_2} with small neighbourhoods of the poles of Z deleted therefrom.

For $k \geq 0$ and any j, m with $1 \leq j \leq a$ and $1 \leq m \leq m_k$, let us define

$$\varphi_{k,j,m}(x) = x^{\mu_{k,j}} (\log x)^{m-1} \text{ for } x \in \mathbb{R}_+^\times;$$

then $\varphi_{k,j,m} = o(\varphi_{k+1,l,n})$ as x tends to 0,

for every j, l, m and n . Further if we define, for $k \geq 0$,

$$R_k(x) = F(x) - \sum_{0 \leq l \leq k} \sum_{1 \leq j \leq a} \sum_{1 \leq m \leq m_l} a_{l,j,m} x^{\mu_{l,j}} (\log x)^{m-1} \quad (2)$$

then the asymptotic expansion (1) for F in $\mathcal{F}$ simply means that $R_k(x) = O(\varphi_{k+1,p,q})$ for every $k \geq 0$ and any p, q as x tends to 0.

The space $\mathcal{F}$ is non-empty, since it contains the constant function $F = 0$ on $\mathbb{R}_+^\times$. Further it is stable under homothety-invariant differential operators; this property of $\mathcal{F}$ corresponds to $\mathcal{Z}$ being stable under multiplication by polynomials in s .

Theorem 1. *There exists a bijective correspondence*

$$M: \mathcal{F} \xrightarrow{\sim} \mathcal{Z} \text{ given by } F \rightarrow MF \text{ where}$$

$$(MF)(s) = \int_0^\infty F(x) x^s d(\log x), \quad (3)$$

defined for $s = \sigma + ti$ in $\mathbb{C}$ with $\sigma = \operatorname{Re}(s) > 0$ has a meromorphic continuation to the whole of $\mathbb{C}$. Conversely, for every Z in $\mathcal{Z}$, the inverse transform

$$(M^{-1}Z)(x) = \frac{1}{2\pi i} \int_{\sigma - \infty i}^{\sigma + \infty i} Z(s) x^{-s} ds \quad (4)$$

gives rise to an element of $\mathcal{F}$, independently of σ for $\sigma > 0$. Moreover, we have the relation

$$b_{l,j,m} = (-1)^{m-1} (m-1)! a_{l,j,m}, \quad (5)$$

for every l, j, m .

The proof of this theorem is exactly the same as in [10] and for the sake of completeness, we shall quickly sketch its proof.

(i) $F \in \mathcal{F} \Rightarrow Z = MF \in \mathcal{Z}$:

Let $s = \sigma + ti$ with $0 < \sigma_1 \leq \sigma \leq \sigma_2 < \infty$. It is easy to see that $x^s F(x)$ is dominated by a φ in $L^1(\mathbb{R}_+^\times, d(\log x))$; therefore $Z(s)$ is holomorphic for $\operatorname{Re} s > 0$, in view of its being defined by an absolutely convergent integral and the integrand being holomorphic. It is clear, for a similar reason that $Z_2(s) = \int_1^\infty F(x) x^s d(\log x)$ is an entire function of s and further bounded in vertical strips. For every $\rho < \lambda_{k+1}$, $R_k(x) = o(x^\rho)$ as x tends to 0 and therefore, for every σ_1 with $-\lambda_{k+1} < \sigma_1 \leq \sigma$ and $-\sigma_1 \leq \rho < \lambda_{k+1}$, the function $x^{\sigma_1} R_k(x)$ is dominated by an element of $L^1((0, 1], d(\log x))$. As a result, $Z_1(s) = \int_0^1 R_k(x) x^s d(\log x)$ is holomorphic in s for $\operatorname{Re}(s) > -\lambda_{k+1}$ and bounded in vertical strips. Splitting up the integral in (3) over $(0, 1]$ and $[1, \infty)$, it is easy to see from (2) that for $\operatorname{Re} s > 0$,

$$Z(s) = \sum_{0 \leq i \leq k} \sum_{1 \leq j \leq a} \sum_{1 \leq m \leq m_i} b_{i,j,m} (s + \mu_{i,j})^{-m} + Z_1(s) + Z_2(s)$$

with $b_{i,j,m}$ satisfying (5). It is now immediate that Z satisfies the first half of condition (B) and also the second half with $P \equiv 1$. For the case of arbitrary P , one has only to invoke the homothety-invariance of $\mathcal{F}$ and it now follows that Z is in $\mathcal{Z}$.

(ii) $Z \in \mathcal{Z} \Rightarrow F = M^{-1} Z \in \mathcal{F}$:

For $0 < \sigma_1 < \sigma_2$, let L denote the boundary (covered anticlockwise) of the rectangle in the complex s -plane with vertices at $\sigma_1 \pm t_0 i, \sigma_2 \pm t_0 i$ for some $t_0 > 0$. By Cauchy's theorem, $\int_L Z(s) x^{-s} ds = 0$ for any $x > 0$. As t_0 tends to ∞ , the contribution to this integral from the "horizontal" sides of L tends to 0, in view of the second part of condition (B) for Z . Thus the integral (4) is defined independently of σ and moreover converges absolutely in view of condition (B). Now, for any $k \geq 0$,

$$x^k \frac{d^k F(x)}{dx^k} = \frac{1}{2\pi i} \int_{\sigma - \infty i}^{\sigma + \infty i} (-1)^k s(s+1) \dots (s+k-1) Z(s) x^{-s} ds$$

and this integral converges absolutely by condition (B). Thus F is in $C^\infty(\mathbb{R}_+^\times)$ and behaves like a Schwartz function at infinity. We then see that $R_k(x)$ defined by (2) satisfies the condition $R_k(x) = o(x^\rho)$ for any ρ with $\lambda_k < \rho < \lambda_{k+1}$ and $k \geq 0$. If L_1 is the boundary traversed anticlockwise of the rectangle in the s -plane with vertices at $\sigma \pm t_0 i, -\rho \pm t_0 i$ for a large enough $t_0 > 0$, one shows by choosing a proper contour and using condition (B) that

$$F(x) = S + \frac{1}{2\pi i} \int_{-\rho - \infty i}^{-\rho + \infty i} Z(s) x^{-s} ds,$$

where S is the sum of the residues of the integrand at the poles $-\mu_{0,1}, \dots, -\mu_{0,a}, -\mu_{1,1}, \dots, -\mu_{1,a}, \dots$, lying inside L_1 . But $F(x) - S$ is precisely $R_k(x)$ and the required o -estimate for $R_k(x)$ follows, in view of (B), by majorising the integral. Now $-sZ(s) = (\mu_{1,p} - (s + \mu_{1,p}))Z(s)$ is in $\mathcal{Z}$ with

$$b'_{i,j,p} = \begin{cases} \mu_{i,j} b_{i,j,p} - b_{i,j,p+1} & 1 \leq p < m_i \\ \mu_{i,j} b_{i,j,p} & p = m_i \end{cases}$$

in place of the earlier $b_{i,j,p}$. Working with $-sZ(s)$ in place of $Z(s)$, we get $xF'(x)$ instead of $F(x)$ and further

$$xF'(x) \approx \sum_{0 \leq i < \infty} \sum_{1 \leq j \leq a} \sum_{1 \leq m \leq m_k} \tilde{a}_{k,i,m} x^{\mu_{k,i}} (\log x)^{m-1}$$

with $\tilde{a}_{k,i,m} = (-1)^{m-1} b'_{k,i,m} / (m-1)!$

$$= \begin{cases} \mu_{k,i} a_{k,i,m} + m a_{k,i,m+1}, & 1 \leq m < m_k \\ \mu_{k,i} a_{k,i,m}, & m = m_k. \end{cases}$$

Thus the asymptotic expansion of F is once termwise differentiable; interaction shows that F is in $\mathcal{F}$.

Remarks : (i) If the condition $\lambda_k \geq 0$ for every k is replaced by a condition of the form $\lambda_k \geq \lambda$ for every k for some fixed real $\lambda \leq 0$, the corresponding $\mathcal{Z}$ -space is obtained from the earlier one by a mere translation of the variable s by λ ; similarly, the corresponding space $\mathcal{F}$ is obtained by multiplying the elements of the original $\mathcal{F}$ -space by x^λ .

(ii) One can also consider more general sequences of the form

$$\mu_{0,1}, \dots, \mu_{0,a_0}, \mu_{1,1}, \dots, \mu_{1,a_1}, \mu_{2,1}, \dots, \mu_{2,a_2}, \dots$$

with $\text{Re}(\mu_{k,j}) = \lambda_k$ for every fixed $k \geq 0$ and for $1 \leq j \leq a_k$. However, arithmetically interesting situations arise when, for example, the sequence of $\mu_{k,j}$'s coincides with the set of poles of a product of Γ -factors, say

$$G(s) = \prod_{1 \leq j \leq a} \Gamma(a_j s + \beta_j)^{m_j} \text{ with } a_j > 0, \text{Re } \beta_j \geq 0 \text{ and}$$

$a_i \beta_j - a_j \beta_i$ not of the form $m a_j - n a_i$ for $m, n \geq 0$ in $\mathbb{Z}$ if $i \neq j$. In such a situation, $\mathcal{Z}$ can be characterised as the space of meromorphic functions Z on $\mathbb{C}$ such that $Z(s)/G(s)$ is entire and for every polynomial P in s , $P(s)Z(s)$ is bounded in vertical strips. In the corresponding space $\mathcal{F}$, we have for every $\kappa > 0$, an involution $F \mapsto WF$ defined by

$$\frac{(M(WF))(s)}{G(s)} = \frac{(MF)(\kappa - s)}{G(\kappa - s)}. \quad (6)$$

(As we shall see later, an analogue of this operator W exists already in the non-archimedean case as well). We shall be interested in obtaining Poisson formulae involved with functional equations containing multiple Γ -factors.

3. Poisson formula of Hecke type

Let $\{\varphi_j(s) = \sum_{n \neq 0} a_n^{(j)} |n|^{-s}; 1 \leq j \leq N\}$ and

$$\{\psi_j^*(s) = \sum_{n \neq 0} b_n^{(j)} |n|^{-s}; 1 \leq j \leq N\}$$

be two sets of N Dirichlet series each converging in some (right) s -halfplane such that if we set, for $1 \leq j \leq N$ and a fixed $A > 0$,

$$\xi_j(s) = A^s G(s) \varphi_j(s), \quad \eta_j^*(s) = A^s G(s) \psi_j^*(s), \quad (7)$$

then we have the functional equations

$$\xi_j(\kappa - s) = \sum_{1 \leq k \leq N} c_{jk} \eta_k^*(s), \quad 1 \leq j \leq N \quad (8)$$

with real c_{jk} . We may assume $\xi_1, \dots, \xi_N$ to be linearly independent over $\mathbf{C}$; then, $(c_{jk})^2$ is the N -rowed identity matrix. We also assume that ξ_k, η_l^* have only finitely many poles in $\mathbf{C}$.

From (6), (7) and (8), we obtain

$$(MWF)(s) A^s \varphi_k(s) = \sum_{1 \leq l \leq N} c_{kl}(MF)(\kappa - s) A^{\kappa-s} \psi_l^*(\kappa - s). \quad (9)$$

In view of the absolute convergence of the Dirichlet series $\varphi_j(s)$ in some right half plane, we see that the integral of the left hand side of (9) from $\sigma - \infty i$ to $\sigma + \infty i$ for any fixed sufficiently large σ is simply

$$2\pi i \sum_{n \neq 0} a_n^{(k)} (WF)(|n|/A).$$

Thus, for large σ_1 , we have

$$\begin{aligned} & \sum_{n \neq 0} a_n^{(k)} (WF)(|n|/A) \\ &= \frac{1}{2\pi i} \int_{\kappa - \sigma - \infty i}^{\kappa - \sigma + \infty i} (M(WF))(\kappa - s) A^{\kappa-s} \varphi_k(\kappa - s) ds \\ &= \frac{1}{2\pi i} \sum_{1 \leq l \leq N} c_{kl} \int_{\kappa - \sigma - \infty i}^{\kappa - \sigma + \infty i} (MF)(s) A^s \psi_l^*(s) ds \\ &= \frac{1}{2\pi i} \sum_{1 \leq l \leq N} c_{kl} \int_{\sigma_1 - \infty i}^{\sigma_1 + \infty i} \frac{(MF)(s)}{G(s)} \eta_l^*(s) ds = S^* \end{aligned}$$

where $S^* = \sum \text{Res} \left(\sum_{1 \leq l \leq N} c_{kl} \frac{(MF)(s)}{G(s)} \eta_l^*(s) \right),$

the sum of the residues of the sum inside at all the poles. Using the absolute convergence of the Dirichlet series $\psi_l^*(s)$, we have

$$\sum_{n \neq 0} a_n^{(k)} (\mathbf{W}F) (|n|/A) = \sum_{1 \leq l \leq N} c_{kl} \sum_{n \neq 0} b_n^{(l)} F(|n|/A) - S^*$$

which, on replacing F by $\mathbf{W}F$ becomes

$$\begin{aligned} \sum_{n \neq 0} a_n^{(k)} F(|n|/A) &= \sum_{1 \leq l \leq N} c_{kl} \sum_{n \neq 0} b_n^{(l)} (\mathbf{W}F) (|n|/A) - \\ &- \sum \text{Res} \left(\frac{(M(\mathbf{W}F))(s)}{G(s)} \xi_k(\kappa - s) \right) \end{aligned} \quad (10)$$

where $\sum \text{Res}$ is the sum of the residues at all poles, as before.

For fixed $y > 0$, let us replace F in (10) by F_y where $F_y(x) = F(xy)$. Then

$$\mathbf{W}(F_y) = y^{-\kappa} (\mathbf{W}F)_{y^{-1}}, \quad (11)$$

where $(\mathbf{W}F)_{y^{-1}}(x) = (\mathbf{W}F)(xy^{-1})$

for every $x > 0$. Therefore

$$M(\mathbf{W}(F_y))(s) = y^{-\kappa} M((\mathbf{W}F)_{y^{-1}})(s) = y^{s-\kappa} (M(\mathbf{W}F))(s)$$

and (10) now becomes

$$\begin{aligned} \sum_{n \neq 0} a_n^{(k)} F(|n|/yA) &= \\ &= y^{-\kappa} \sum_{1 \leq l \leq N} c_{kl} \sum_{n \neq 0} b_n^{(l)} (\mathbf{W}F) (|n|/(Ay)) - \\ &- \sum \text{Res} \left(\frac{(M(\mathbf{W}F))(s)}{G(s)} y^{s-\kappa} \xi_k(\kappa - s) \right) \end{aligned} \quad (10)'$$

$$\begin{aligned} &= y^{-\kappa} \sum_{1 \leq l \leq N} c_{kl} \sum_{n \neq 0} b_n^{(l)} (\mathbf{W}F) (|n|/(Ay)) + \\ &+ \sum \text{Res} \left(y^{-s} \frac{(MF)(s)}{G(s)} \xi_k(s) \right). \end{aligned} \quad (10)''$$

The second term in (10)'' is a 'residual function' in the sense of Bochner [3] and (10)'' gives rise to a "generalised modular relation",

For the sake of simplicity, we assume that no $\xi_k(s)$ has a pole on the line $\operatorname{Re}(s) = \kappa/2$. Let $u_1, \dots, u_p$ be the poles of $\xi_k(s)$. Then we may rewrite (10) as

$$\begin{aligned} & \sum_{n \neq 0} a_n^{(k)} F(|n|/A) + \sum_{\operatorname{Re} u_j > \kappa/2} \operatorname{Res}_{s=u_j} \frac{(M(WF))(s)}{G(s)} \xi_k(\kappa - s) \\ &= \sum_{1 \leq l \leq N} c_{kl} \sum_{n \neq 0} b_n^{(l)}(WF)(|n|/A) - \sum_{\operatorname{Re} u_j < \kappa/2} \operatorname{Res}_{s=u_j} \frac{(M(WF))(s)}{G(s)} \xi_k(\kappa - s). \end{aligned} \quad (12)$$

The second term on the left hand side of (12) is the same as

$$- \sum_{\operatorname{Re} u_j > \kappa/2} \operatorname{Res}_{s=\kappa-u_j} \frac{MF(s)}{G(s)} \xi_k(s) = - \sum_{\operatorname{Re} u_j < \kappa/2} \operatorname{Res}_{s=u_j} \frac{(MF)(s)}{G(s)} \xi_k(s)$$

while the one on its right hand side is equal to

$$- \sum_{1 \leq l \leq N} c_{kl} \sum_{\operatorname{Re} u_j < \kappa/2} \operatorname{Res}_{s=u_j} \frac{(M(WF))(s)}{G(s)} \eta_l^*(s)$$

Thus we have a Poisson formula of Hecke type given by the following *Theorem 2*. For any F whose Mellin transform MF is such that $(MF)(s)/G(s)$ is entire and $P(s)MF(s)$ is bounded in vertical strips for every polynomial P and for $\xi_1(s), \dots, \xi_N(s), \eta_1^*(s), \dots, \eta_N^*(s)$ satisfying functional equations (8), we have

$$\begin{aligned} & \sum_{n \neq 0} a_n^{(k)} F(|n|/A) - \sum_{\operatorname{Re} u_j < \kappa/2} \operatorname{Res}_{s=u_j} \frac{MF(s)}{G(s)} \xi_k(s) \\ &= \sum_{1 \leq l \leq N} c_{kl} \left(\sum_{n \neq 0} b_n^{(l)}(WF)(|n|/A) \right. \\ & \quad \left. - \sum_{\operatorname{Re} u_j < \kappa/2} \operatorname{Res}_{s=u_j} \frac{(M(WF))(s)}{G(s)} \eta_l^*(s) \right). \end{aligned} \quad (13)$$

4. Known cases of Poisson formulae above

Formula (13) generalises some well-known relations of a similar nature due to Hecke, Maass and Yamazaki and may be called a "generalized modular relation".

(1) First let $a = 1$, $N = 1$, $\mu_{k,1} = k$ for $k \geq 0$, $G(s) = \Gamma(s)$,

$$\xi_1(s) = \xi(s) = \gamma \xi(\kappa - s) = \eta_1^*(s), \quad c_{11} = \gamma, \quad \varphi_1(s) = \varphi(s) = \sum_{n=1}^{\infty} a_n n^{-s},$$

$A = \lambda/2\pi$; let $s = 0$, κ be the only poles (and indeed of order 1) of $\xi(s)$. Then

$$\begin{aligned}
 \operatorname{Res}_{s=0} \frac{MF(s)}{G(s)} \xi_1(s) &= b_{0,1,1} \operatorname{Res}_{s=0} \xi(s) && \text{by Theorem 1} \\
 &= a_{0,1,1} \operatorname{Res}_{s=0} \xi(s) && \text{by Theorem 1} \\
 &= F(0) \times (-\gamma \operatorname{Res}_{s=\kappa} A^s F(s) \varphi(s)) \\
 &= -\gamma F(0) A^\kappa \Gamma(\kappa) \operatorname{Res}_{s=\kappa} \varphi(s) \\
 &= -CF(0), \text{ say.}
 \end{aligned}$$

Replacing F by WF in the argument above, we obtain

$$\begin{aligned}
 \operatorname{Res}_{s=0} \frac{(M(WF))(s)}{G(s)} \eta_1^*(s) &= \operatorname{Res}_{s=0} \frac{(M(WF))(s)}{G(s)} \xi(s) \\
 &= -C(WF)(0).
 \end{aligned}$$

Thus we get from (13) the Poisson formula due to Yamazaki [19] in the form derived by Igusa [10]: namely, for every F in $\mathcal{F}$ (with $G(s) = \Gamma(s)$)

$$CF(0) + \sum_{n \geq 1} a_n F(2\pi n/\lambda) = \gamma (C(WF)(0) + \sum_{n \geq 1} a_n (WF)(2\pi n/\lambda)). \quad (14)$$

(2) Let us take the same situation in (1) above and in particular,

$$\varphi(s) = \sum_{1 \leq n < \infty} \tau(n) n^{-s}$$

where $\tau(n)$ is Ramanujan's function; then $\kappa = 12$, $A = 1/2\pi$, $\xi(s)$ is entire and for every F in the space $\mathcal{F}$, we have

$$\sum_{1 \leq n < \infty} \tau(n) F(2\pi n) = \sum_{1 \leq n < \infty} \tau(n) (WF)(2\pi n). \quad (15)$$

For $F(x) = \exp(-xy/2\pi)$ with fixed $y > 0$, we obtain in view of (11) and (15), formula (4) of [4]. On the other hand, if we take $F(x) = \exp(-s\sqrt{x})/\sqrt{x}$ for fixed $s > 0$, then we obtain *formally* from (15), formula (7) of [4]. One has to note that in the situation of (1) above, we can define for any F in $\mathcal{F}$, the function WF also by

$$(WF)(x) = \int_0^\infty (xt)^{-\nu/2} J_\nu(2\sqrt{xt}) F(t) t^\nu dt, \quad (16)$$

with $\nu = \kappa - 1$. (See [10]), where J_ν is the usual Bessel function of order ν . However, the function $x \mapsto [\exp(-s\sqrt{x})]/\sqrt{x}$ is *not* in the space $\mathcal{F}$ above. If we approximate to it by a sequence $\{F_n\}$ from $C_c^\infty(\mathbb{R}_+^\times)$ (and indeed therefore with F_n in $\mathcal{F}$) and note that $\{MF_n\}$ converges to the (usual) Mellin transform MF

of F while $\{WF_n\}$ converges to WF (defined in the same way as in (16)), the validity of formula (7) of [4] can be deduced from above, in view of the two sides of this formula being absolutely convergent series. Again, for *integral* $\rho > 0$, formula (5) of [4] can be derived from (15), by taking for F , the function in $\mathcal{F}$ given by

$$F(t) = \begin{cases} \frac{1}{\Gamma(\rho+1)} (1-tx)^\rho & \text{for } 0 < t \leq x^{-1} \\ 0 & \text{for } t > x^{-1}, \end{cases}$$

with fixed $x > 0$. On the other hand, for getting the same formula for non-integral $\rho > 0$, a more refined argument as above has to be used; the case $\rho = 0$ also needs a more careful argument.

(3) When $\varphi(s)$ is the Dedekind zeta function associated with an imaginary quadratic field of discriminant d over the field of rational numbers and $A = \sqrt{|d|}/2\pi$ in (1) above, then we get the Poisson formula due to Hecke [8].

(4) The relation in Kubota [11] corresponding to our relation (9) above is, in his own notation,

$$M\Psi_b(s) Z_{a,b}(s) = M\tilde{\Psi}_{-b}(2n-2-s) Z_{-a,-b}(2n-2-s),$$

and proceeding exactly as above, one can derive his Poisson formula

$$\sum_{(m)} c'_m \Phi_b(m^{1/n}) = \sum_{(m)} c'_m \Phi_b^*(m^{1/n})$$

and also his relation

$$\begin{aligned} -AM(\Psi_0, 2n-2) + \sum_{(m)} c'_m \Phi_0(m^{1/n}) \\ = -AM(\tilde{\Psi}_0, 2n-2) + \sum_{(m)} c'_m \Phi_0^*(m^{1/n}) \end{aligned}$$

(see pages 187-188 in [11]).

(5) Let now $a = 2$, $\mu_{k,1} = 2k + ir$, $\mu_{k,2} = 2k - ir$ for $k = 0, 1, 2, \dots$ and a fixed $r \geq 0$, $m_k = 1$ or 2 according as $r \neq 0$ or $r = 0$,

$$A = \lambda/\pi, \varphi_k(s) = \psi_k^*(s) = \sum_{n \neq 0} a_n^{(k)} |n|^{-s} \quad (1 \leq k \leq N) \quad \text{with } a_n^{(k)} = 0$$

unless $n \equiv b_k \pmod{q}$ for fixed integers $q \geq 1$ and

$$b_1, \dots, b_N, \kappa = 1, G(s) = \Gamma\left(\frac{s-ir}{2}\right) \Gamma\left(\frac{s+ir}{2}\right)$$

and $\xi_k(s) = \eta_k^*(s) = A^s G(s) \varphi_k(s)$ for $1 \leq k \leq N$.

$$\text{Let } \varphi_k(s) = \begin{cases} \frac{a_k}{s-1-ir} + \frac{\beta_k}{s-1+ir} & (r \neq 0) \\ \frac{a_k}{s-1} + \frac{\beta_k}{(s-1)^2} & (r = 0) \end{cases}$$

be entire (and of finite genus) for $1 \leq k \leq N$. Equivalently ([13]), if now C denotes Euler's constant and

$$M_0 = \frac{\sqrt{\lambda}}{4} \Gamma\left(\frac{1}{2} + ir\right) (\lambda/\pi)^{\frac{1}{2} + ir},$$

$$\text{Let } \xi_k(s) = \begin{cases} \frac{4M_0 a_k}{s-1-ir} + \frac{4\bar{M}_0 \beta_k}{s-1+ir} - \frac{4M_0 \rho_k}{s+ir} - \frac{4\bar{M}_0 \sigma_k}{s-ir} & (r \neq 0) \\ \frac{4M_0}{s-1} (a_k + \beta_k (\log(\lambda/4\pi) - C)) + \frac{4M_0 \beta_k}{(s-1)^2} - \frac{4M_0}{s} \times \\ \times (\rho_k + \sigma_k) (\log(\lambda/4\pi) - C) + \frac{4M_0 \sigma_k}{s^2} & (r = 0) \end{cases}$$

be entire in s (and of finite genus) with

$$\rho_k = \sum_{1 \leq i \leq N} c_{ki} a_i$$

$$\text{and } \sigma_k = \sum_{1 \leq i \leq N} c_{ki} \beta_i, \text{ for } 1 \leq k \leq N.$$

Then for F in the corresponding $\mathcal{F}$ -space, we have as t tends to 0, the asymptotic expansion

$$F(t) \approx \begin{cases} a_{0,1,1} t^{ir} + a_{0,2,1} t^{-ir} + \sum_{1 \leq k < \infty} (a_{k,1,1} t^{2k+ir} + a_{k,2,1} t^{2k-ir}) & (r \neq 0) \\ a_{0,1,1} + a_{0,1,2} \log t + \sum_{1 \leq k < \infty} (a_{k,1,1} + a_{k,1,2} \log t) t^{2k} & (r = 0). \end{cases}$$

By Theorem 1 above,

$$(MF)(s) = \begin{cases} \frac{a_{0,1,1}}{s+ir} + \frac{a_{0,2,1}}{s-ir} & (r \neq 0) \\ \frac{a_{0,1,1}}{s} - \frac{a_{0,1,2}}{s^2} & (r = 0). \end{cases} \text{ is regular at } s = \pm ir. \quad (17)$$

If $r \neq 0$, the only poles of $\xi_k(s)$ to the left of the line $\text{Re } s = \frac{1}{2}$ are at $s = \pm ir$ and of order 1. Then

$$\begin{aligned} -\text{Res}_{s=ir} \frac{(MF)(s)}{G(s)} \xi_k(s) &= - \frac{(MF)(s)}{\Gamma\left(\frac{s+ir}{2}\right) \Gamma\left(\frac{s-ir}{2}\right)} \Big|_{s=ir} \times \text{Res}_{s=ir} \xi_k(s) \\ &= 2\bar{M}_0 a_{0,2,1}(F) \sigma_k / \Gamma(ir), \end{aligned}$$

writing $a_{0,j,m}(F)$ instead of $a_{j,m}$ to emphasise the dependence on F . Similarly we have

$$-\text{Res}_{s=-ir} \frac{(MF)(s)}{G(s)} \xi_k(s) = 2M_0 a_{0,1,1}(F) \rho_k / \Gamma(-ir).$$

In order to find

$$-\operatorname{Res}_{s=\pm ir} \frac{(M(WF))(s)}{G(s)} \eta_i^*(s),$$

we have only to work with WF instead of F in the arguments above.

If $r = 0$, the only pole of $\xi_k(s)$ to the left of the line $\operatorname{Re} s = 1$ is at $s = 0$ and indeed it is of order 2. Using (17) and the expansion $1/\Gamma^2(s/2) = (s^2/4)(1 + Cs + \dots)$ at $s = 0$, we obtain

$$\operatorname{Res}_{s=0} \frac{(MF)(s)}{\Gamma^2(s/2)} \xi_k(s) = a_{0,1,1}(F) M_0 \sigma_k + a_{0,1,2}(F) M_0 (\rho_k + \sigma_k (\log(\lambda/4\pi) - 2C)).$$

The residue of $(M(WF))(s) \xi_k(s)/\Gamma^2(s/2)$ at $s = 0$ is obtained by arguing with WF in lieu of F above. Thus, from (13), we have, for any F in the space $\mathcal{F}$, the Poisson formula

$$\begin{aligned} & \sum_{n \neq 0} a_n^{(k)} F(|n|/A) + \\ & \left\{ 2M_0 \rho_k a_{0,1,1}(F)/\Gamma(-ir) + 2\bar{M}_0 \sigma_k a_{0,2,1}(F)/\Gamma(ir) \right. \\ & \left. - M_0 \sigma_k a_{0,1,1}(F) - M_0 (\rho_k + \sigma_k (\log(\lambda/4\pi) - 2C)) a_{0,1,2}(F) \right\} \quad (18) \\ & = \sum_{1 \leq i \leq N} c_{ki} \left[\sum_{n \neq 0} a_n^{(i)}(WF)(|n|/A) \right. \\ & \left. + \left\{ 2M_0 \rho_i a_{0,1,1}(WF)/\Gamma(-ir) + 2\bar{M}_0 \sigma_i a_{0,2,1}(WF)/\Gamma(ir) \right. \right. \\ & \left. \left. - M_0 \sigma_i a_{0,1,1}(WF) - M_0 (\rho_i + \sigma_i (\log(\lambda/4\pi) - 2C)) a_{0,1,2}(WF) \right\} \right] \end{aligned}$$

according as $r \neq 0$ or $r = 0$. The roles of $F(0)$ and $(WF)(0)$ in (14) are played now by the coefficients $a_{0,j,m}(F)$ and $a_{0,j,m}(WF)$ in their asymptotic expansions. If we take $F(t) = 4\sqrt{y} K_{it}(2ty)$ for $t > 0$ with a fixed $y > 0$, then

$$a_{0,1,1}(F) = 2y^{\frac{1}{2}+ir} \Gamma(-ir), \quad a_{0,2,1}(F) = 2y^{\frac{1}{2}-ir} \Gamma(ir), \quad (r \neq 0)$$

$$a_{0,1,1}(F) = -4\sqrt{y}(C + \log y), \quad a_{0,1,2}(F) = -4\sqrt{y}, \quad (r = 0)$$

$$MF(s) = y^{\frac{1}{2}-s} \Gamma\left(\frac{s+ir}{2}\right) \Gamma\left(\frac{s-ir}{2}\right).$$

Similar formulae for WF are valid; we have only to replace y by $1/y$ and F by WF in the formulae above. With this specialisation, formula (18) is the same as the "automorphic" relation

$$F_k^*(1/y) = \sum_{1 \leq i \leq N} c_{ki} F_i^*(y)$$

of Maass ([13], p. 152).

(6) If $k = \mathbf{Q}(\sqrt{d})$ is a real quadratic field of discriminant d over the field $\mathbf{Q}$ of rational numbers, then the Dedekind zeta function $\zeta_k(s)$ satisfies the functional equation $(\pi/\sqrt{d})^{-s} \Gamma^2(s/2) \zeta_k(s) = \xi(s) = \xi(1-s)$. The function $\zeta_k(s)$ has a pole only at $s=1$ and the residue at this (simple) pole is $2h(\log \epsilon)/\sqrt{d}$ where h is the class number of K and ϵ is the fundamental unit in K . Moreover, $\zeta_k(0) = 0$ and $\zeta'_k(0) = -h(\log \epsilon)/2$. If we take

$$\varphi(s) = (\pi/\sqrt{d})^{-2s} \zeta_k(2s) = \sum_{n \geq 1} a(n) \lambda_n^{-s}$$

with $\lambda_n = \pi^2 n^2/d$, then $\varphi(s) \Gamma^2(s) = \varphi(1/2-s) \Gamma^2(1/2-s)$. From Theorem 1 of Berndt [2], we get (on noting that his $E_2(y)$ is just $2K_0(2\sqrt{y})$) the following Poisson formula,

$$2 \sum_{1 \leq n < \infty} a(n) K_0(2\pi n y/d) = 2y^{-1} \sum_{1 \leq n < \infty} a(n) K_0(2\pi n/(y\sqrt{d})) + P(y^2)$$

where K_0 is the usual Bessel function. Here $P(y^2) = \text{sum of the residues of } y^{-2s} \Gamma^2(s) \varphi(s)$ is seen to be equal to $h(\log \epsilon)/y + 2\zeta'_k(0) = h(\log \epsilon)(y^{-1} - 1)$. Now this formula reduces to a special case of the Poisson formula of Maass mentioned in (5) above.

5. A Poisson formula associated with a generalised Γ -function

In this section, we derive a Poisson formula for a situation involving Dirichlet series with functional equations containing a generalised Γ -function $\Gamma(s; a, \beta)$ introduced by Maass ([14], [15]). This function $\Gamma(s; a, \beta)$ is not, in general, a product of usual Γ -functions although, however, $\Gamma(s; a, \beta)/(\Gamma(s) \Gamma(s+1-a-\beta))$ is an entire function of s (with finite genus). The question is one of defining correctly the W -operator geared to the Poisson formula in this case; one has been guided here by the definition of a "Hankel transform" through the two-component Mellin transform for L^2 -functions on $\mathbf{R} \setminus \{0\}$ defined in ([16], Theorem 10) (see also [17]).

First we recall the definition of the Whittaker functions $W_{l,m}(y)$ for $l, m \in \mathbf{C}$ and $y > 0$ as the "unique" solution $W(y)$ of the differential equation

$$4y^2 \frac{d^2 W}{dy^2} + (1 - 4m^2 + 4ly - y^2) W(y) = 0,$$

with the asymptotic behaviour

$$W(y) \sim \exp(-y/2) y^l \left\{ 1 + \sum_{1 \leq n < \infty} \frac{1}{y^n n!} \prod_{r=1}^n (m^2 - (l + \frac{1}{2} - r)^2) \right\}$$

as y tends to ∞ . Let

$$W(y; a, \beta) = y^{-(a+\beta)/2} W_{(a-\beta)/2, (a+\beta-1)/2}(2y),$$

for $y > 0$ and

$$\Gamma(s; a, \beta) = \int_0^\infty W(y; a, \beta) y^{s-1} dy,$$

for $\operatorname{Re} s > K_0 = \max(0, |\operatorname{Re}(a + \beta)| - 1)$. Then $\Gamma(s; a, \beta)$ is regular for $\operatorname{Re} s > K_0$, noting that $W(y; a, \beta) = O(y^{-K})$ for $K > K_0$ as y tends to 0 and further it has a meromorphic continuation to the entire s -plane satisfying the condition

$$\Gamma(s; a, \beta) = 2^{(a-\beta)/2} \frac{\Gamma(s) \Gamma(s+1-a-\beta)}{\Gamma(s+1-a)} F(\beta, 1-a, s+1-a; \frac{1}{2})$$

where $F(a, b, c; z) = 1 + \sum_{1 \leq n < \infty} \frac{a(a+1) \dots (a+n-1) b(b+1) \dots (b+n-1)}{c(c+1) \dots (c+n-1) \cdot 1 \cdot 2 \dots n} z^n$

is the hypergeometric function. It is also known that for every polynomial P $P(s) \Gamma(s; a, \beta)$ is bounded in vertical strips in the s -plane. If we set

$$M(s; a, \beta) = \begin{pmatrix} \Gamma(s; a, \beta) & \Gamma(s; \beta, a) \\ \Gamma(s+1; a, \beta) & \Gamma(s+1; \beta, a) \end{pmatrix}$$

then we know from Maass [15] that its determinant $D(s)$ is $-2 \Gamma(s) \times \Gamma(s+1-a-\beta)$; further, the entries of the inverse matrix are entire functions of s . Let us write q for $a + \beta$ in the sequel and assume that $q \neq 1$.

Before we go on to the Poisson formula, let us introduce the Dirichlet series whose functional equations involve the function $\Gamma(s; a, \beta)$.

$$\text{Let } \varphi(s) = \sum_{1 \leq n < \infty} a_n n^{-s} \text{ and } \psi(s) = \sum_{-\infty < n < 0} a_n |n|^{-s},$$

be two Dirichlet series converging absolutely in some s -half plane and let, for some $\lambda > 0$,

$$\begin{aligned} \xi(s) &= (\lambda/2\pi)^s (\Gamma(s; a, \beta) \varphi(s) + \Gamma(s; \beta, a) \psi(s)), \\ \eta(s) + \lambda((a - \beta)/4\pi) \xi(s) &= (\lambda/2\pi)^{s+1} (\Gamma(s+1; a, \beta) \varphi(s) - \\ &\quad - \Gamma(s+1; \beta, a) \psi(s)) \end{aligned} \quad (19)$$

satisfy the functional equations

$$\xi(q-s) = \gamma \xi(s), \eta(q-s) = -\gamma \eta(s), \quad (20)$$

with fixed $\gamma = \pm 1$. Further, let us assume that

$$\xi(s), \eta_1(s) = \eta(s) + \lambda((a - \beta)/4\pi) \xi(s)$$

have poles at most at $s = 0, 1, q-1, q$ with principal parts as given by the conditions:

$$\xi(s) - \frac{a_0}{s(s+1-q)} - \frac{\gamma a_0}{(q-s)(1-s)} + \frac{b_0}{s} + \frac{b_0}{q-s}$$

is entire, and

$$\eta_1(s) - \lambda \frac{a - \beta}{2\pi} \left(\frac{b_0}{s - q} + \gamma \frac{a_0}{(s - q)(s - 1)} \right) \quad (21)$$

is entire with suitable constants a_0, b_0 . The relations (19) can be inverted to read

$$\begin{aligned} \varphi(s) &= -\frac{\Gamma(s+1; \beta, a)}{D(s)} (2\pi/\lambda)^s \xi(s) - \frac{\Gamma(s; \beta, a)}{D(s)} (2\pi/\lambda)^{s+1} \eta_1(s) \\ \psi(s) &= -\frac{\Gamma(s+1; a, \beta)}{D(s)} (2\pi/\lambda)^s \xi(s) + \frac{\Gamma(s; a, \beta)}{D(s)} (2\pi/\lambda)^{s+1} \eta_1(s). \end{aligned} \quad (19)'$$

The only poles of φ and ψ are at $s = 1, q$. Now functional equations (20) go over into

$$\begin{aligned} (\lambda/2\pi)^s M(s; a, \beta) \begin{pmatrix} \varphi(s) \\ \psi(s) \end{pmatrix} &= (\lambda/2\pi)^{a-s} \begin{pmatrix} 1 & 0 \\ a- & \beta-1 \end{pmatrix} M(q-s; a, \beta) \\ &\quad \begin{pmatrix} \varphi(q-s) \\ \psi(q-s) \end{pmatrix}. \end{aligned} \quad (20)'$$

From (19)' and (21), we obtain

$$\begin{aligned} \varphi(0) &= \frac{1-q}{2} w(1; \beta, a) \left(\frac{a_0}{1-q} - b_0 \right) + \frac{\pi}{\lambda} w(0; \beta, a) \eta_1(0), \\ \varphi(q-1) &= -\left(\frac{2\pi}{\lambda} \right)^{a-1} w(q; \beta, a) \frac{a_0}{2} + \frac{1}{2} \left(\frac{2\pi}{\lambda} \right)^a w(q-1; \beta, a) \eta_1(q-1), \\ \psi(0) &= \frac{1-q}{2} w(1; a, \beta) \left(\frac{a_0}{1-q} - b_0 \right) - \frac{\pi}{\lambda} w(0; a, \beta) \eta_1(0) \times \\ \psi(q-1) &= -\left(\frac{2\pi}{\lambda} \right)^{a-1} w(q; a, \beta) \frac{a_0}{2} - \frac{1}{2} \left(\frac{2\pi}{\lambda} \right)^a w(q-1; a, \beta) \\ &\quad \times \eta_1(q-1) \end{aligned} \quad (22)$$

where $w(s; a, \beta) = -2\Gamma(s; a, \beta)/D(s)$ and $w(s; \beta, a)$ is defined similarly.

Let us now consider the space $\mathfrak{F}$ of C^∞ functions F on $\mathbb{R}_+^\times$ which behave like Schwartz functions at infinity and which have the termwise differentiable asymptotic expansion

$$F(x) \approx \sum_{0 \leq n < \infty} a_n(F) x^n + x^{1-a} \sum_{0 \leq n < \infty} \beta_n(F) x^n,$$

as x tends to 0. From Theorem 1, we, know that $MF(s)/(\Gamma(s)\Gamma(s+1-q))$ is entire and further, for every polynomial P in s , $P(s)MF(s)$ is bounded in vertical strips. We assume that $0 < \operatorname{Re} q < 1$ in the sequel.

Let F_1, F_2 in $\mathfrak{F}$ above satisfy the conditions:

- (i) for every $\epsilon > 0$, $MF_j(s) = O(\exp(-\epsilon|t|))$ as $s = \sigma + ti$ tends to infinity in vertical strips B_{a_j, σ_j} , for $j = 1, 2$, and
- (ii) $(MF_1(s)MF_2(s))M(s; a, \beta)^{-1} = (G_1(s)G_2(s))$ with entire G_1, G_2 . For such a pair F_1, F_2 , we define WF_1, WF_2 through the functional equation

(23)

$$\begin{aligned}
 & ((M(WF_1))(s) M(WF_2)(s)) M(s; a, \beta)^{-1} \\
 &= (G_1(q-s) G_2(q-s)) \begin{pmatrix} 1 & 0 \\ \alpha- & \beta-1 \end{pmatrix}.
 \end{aligned} \tag{24}$$

From (23) and (24), it is clear that the left hand side of (24) is entire. Further for every polynomial P in s , $P(s) (M(WF_i))(s)$ is bounded in vertical strips, for $i=1, 2$ (condition (i) in (23) has been imposed in order to ensure this and if perhaps

$$M(q-s; a, \beta)^{-1} \begin{pmatrix} 1 & 0 \\ \alpha- & \beta-1 \end{pmatrix} M(s; a, \beta)$$

consists of functions bounded at infinity in vertical strips, then condition (i) can be waived). Thus, as x tends to 0, WF_i has the termwise differentiable asymptotic expansion

$$(WF_i)(x) \approx \sum_{0 \leq n < \infty} a_n(WF_i) x^n + x^{1-a} \sum_{0 \leq n < \infty} \beta_n(WF_i) x^n.$$

Replacing s by $q-s$ in (20)' and then multiplying out relations (20)' and (24) we obtain

$$\begin{aligned}
 & \lambda/(2\pi)^{a-s} ((M(WF_1))(q-s) \varphi(q-s) + (M(WF_2))(q-s) \psi(q-s)) \\
 &= \gamma (\lambda/2\pi)^s (MF_1(s) \varphi(s) + MF_2(s) \psi(s)).
 \end{aligned} \tag{25}$$

In order to get the Poisson formula, we proceed as in the proof of Theorem 2. In view of the absolute convergence of φ, ψ for sufficiently large $\text{Re } s$, we see that the integral of the right hand side of (25) from $\sigma - \infty i$ to $\sigma + \infty i$ for σ large enough, is simply

$$2\pi i \gamma \left(\sum_{n > 0} a_n F_1(2\pi n/\lambda) \right) + \sum_{n < 0} a_n F_2(2\pi |n|/\lambda),$$

which, by (25), is therefore equal to

$$\begin{aligned}
 & \int_{\sigma - \infty i}^{\sigma + \infty i} (\lambda/2\pi)^{a-s} ((M(WF_1))(q-s) \varphi(q-s) + (M(WF_2))(q-s) \psi(q-s)) ds \\
 &= \int_{q - \sigma - \infty i}^{q - \sigma + \infty i} (\lambda/2\pi)^s [(M(WF_1))(s) \varphi(s) + (M(WF_2))(s) \psi(s)] ds \\
 &= \int_{\sigma_1 - \infty i}^{\sigma_1 + \infty i} (\lambda/2\pi)^s (M(WF_1)(s) \varphi(s) + (M(WF_2)(s) \psi(s)) ds - 2\pi i S^*,
 \end{aligned}$$

for σ_1 large enough, with S denoting the sum of the residues of the integrand at all the poles encountered when the line of integration is shifted from

$\operatorname{Re}(s) = q - \sigma$ to $\operatorname{Re}(s) = \sigma_1$ far to the right. Because of the absolute convergence of the Dirichlet series $\varphi(s), \psi(s)$ again, we obtain the relation

$$\left(\sum_{n>0} a_n F_1(2\pi n/\lambda) + \sum_{n<0} a_n F_2(2\pi |n|/\lambda) \right) = \sum_{n>0} a_n (WF_1)(2\pi n/\lambda) \\ + \sum_{n<0} a_n (WF_2)(2\pi |n|/\lambda) - S^*.$$

This can be rewritten as in the proof of Theorem 2 as

$$\sum_{n>0} a_n F_1(2\pi n/\lambda) + \sum_{n<0} a_n F_2(2\pi |n|/\lambda) - \sum_{\operatorname{Re} u_j < q/2} \operatorname{Res}((\lambda/2\pi)^s MF_1(s) \varphi(s) \\ + MF_2(s) \psi(s)) \\ = \gamma \sum_{n>0} a_n (WF_1)(2\pi n/\lambda) + \gamma \sum_{n<0} a_n (WF_2)(2\pi |n|/\lambda) - \sum_{\operatorname{Re} u_j < q/2} \operatorname{Res}((\lambda/2\pi)^s \times \\ \times ((M(WF_1))(s) \varphi(s) + (M(WF_2))(s) \psi(s)), \quad (26)$$

where the third summations on both sides are over the residues at all the poles u_j satisfying the condition stated. In view of our assumption that $0 < \operatorname{Re} q < 1$, $s = 0$, $q - 1$ are the only poles involved in these summations. The residue of $(\lambda/2\pi)^s (MF_1(s) \varphi(s) + MF_2(s) \psi(s))$ at 0 is seen to be equal to

$$((1 - q)/2) (a_0/(1 - q) - b_0) (a_0(F_1) w(1, \beta, a) + a_0(F_2) w(1, a, \beta)) - \\ - (\pi/\lambda) \{w(0, \beta, a) a_0(F_1) - w(0, a, \beta) a_0(F_2)\} \gamma_1(0),$$

in view of (22). The expression inside the curly brackets is 0, since it is essentially the residue at $s = 0$ of the right hand side of (25) while the left hand side of (25) is regular at $s = 0$. Similarly, the residue at $q - 1$ of

$$(\lambda/2\pi)^s (MF_1(s) \varphi(s) + MF_2(s) \psi(s))$$

is seen to be just

$$- a_0 (\beta_0(F_1) w(q; \beta, a) + \beta_0(F_2) w(q; a, \beta))/2.$$

The residues involved in the third summation on the right hand side of (25) are computed just as above, replacing F_i by WF_i everywhere. Thus formula (26) leads to

Theorem 3. For C^∞ functions F_1, F_2 on $\mathbb{R}_+^\times$ behaving like Schwartz functions at infinity and satisfying (23) and Dirichlet series $\varphi(s), \psi(s)$ for which the functional equations (20) hold, we have, for $0 < \operatorname{Re} q < 1$, the Poisson formula

$$\sum_{n>0} a_n F_1(2\pi n/\lambda) + \sum_{n<0} a_n F_2(2\pi |n|/\lambda) - \frac{1}{2} a_0 (a_0(F_1) w(1; \beta, a) \\ + a_0(F_2) w(1; a, \beta) - \beta_0(F_1) w(q; \beta, a) - \beta_0(F_2) w(q; a, \beta)) - \\ - b_0(q - 1) (a_0(F_1) w(1, \beta, a) + a_0(F_2) w(1, a, \beta))$$

$$\begin{aligned}
&= \gamma \left[\sum_{n \geq 0} a_n(\mathbf{W}F_1) (2\pi n/\lambda) + \sum_{n \leq 0} a_n(\mathbf{W}F_2) (2\pi |n|/\lambda) - \right. \\
&\quad - \frac{1}{2} a_0(a_0(\mathbf{W}F_1) w(1; \beta, a) + a_0(\mathbf{W}F_2) w(1; a, \beta) \\
&\quad - \beta_0(\mathbf{W}F_1) w(q; \beta, a) - \beta_0(\mathbf{W}F_2) w(q; a, \beta) \\
&\quad \left. - \frac{1}{2} b_0(q-1)(a_0(\mathbf{W}F_1) w(q; \beta, a) + a_0(\mathbf{W}F_2) w(1; a, \beta)) \right],
\end{aligned}$$

where $a_0(F_i)$, $\beta_0(F_i)$, etc., are coefficients in the asymptotic expansions at 0.

Note. The values of $w(0; a, \beta)$, etc., are given explicitly as follows:

$$\begin{aligned}
w(0; a, \beta) &= 2^{a/2}/\Gamma(1-a), \quad w(1; a, \beta) = \Gamma(1; a, \beta)/\Gamma(2-q), \\
w(0; \beta, a) &= 2^{a/2}/\Gamma(1-\beta), \quad w(1; \beta, a) = \Gamma(1; \beta, a)/\Gamma(2-q), \\
w(q-1; a, \beta) &= 2^{1-a/2}/\Gamma(\beta), \quad w(q; a, \beta) = \Gamma(q; a, \beta)/\Gamma(q), \\
w(q-1; \beta, a) &= 2^{1-a/2}/\Gamma(a), \quad w(q; \beta, a) = \Gamma(q; \beta, a)/\Gamma(q).
\end{aligned}$$

Let now $F_1(x) = W(xy; a, \beta)$, $F_2(x) = W(xy; \beta, a)$ for fixed $y > 0$. Then by (11), $(\mathbf{W}F_1)(x) = W(xy^{-1}; a, \beta)y^{-a}$, $(\mathbf{W}F_2)(x) = W(xy^{-1}; \beta, a)y^{-a}$. Further

$$\begin{aligned}
a_0(F_1) \Gamma(1-a) &= a_0(F_2) \Gamma(1-\beta) = 2^{a/2} \Gamma(1-q), \\
\beta_0(F_1) \Gamma(\beta) &= \beta_0(F_2) \Gamma(a) = 2^{1-a/2} y^{1-a} \Gamma(q-1), \\
a_0(\mathbf{W}F_1) \Gamma(1-a) &= a_0(\mathbf{W}F_2) \Gamma(1-\beta) = 2^{a/2} y^{-a} \Gamma(1-q), \\
\beta_0(\mathbf{W}F_1) \Gamma(\beta) &= \beta_0(\mathbf{W}F_2) \Gamma(a) = 2^{1-a/2} y^{a-1} \cdot y^{-a} \Gamma(q-1), \\
\Gamma(a) \Gamma(q; \beta, a) + \Gamma(\beta) \Gamma(q; a, \beta) \\
&= 2^{a/2} \Gamma(q) \{B_{1/2}(\beta, a) + B_{1/2}(a, \beta)\} = 2^{a/2} \Gamma(q) B_1(a, \beta) \\
&= 2^{a/2} \Gamma(a) \Gamma(\beta) \quad \text{where } B_x(a, \beta) = \int_0^x t^{a-1} (1-t)^{\beta-1} dt,
\end{aligned}$$

$$\text{and} \quad \Gamma(1-\beta) \Gamma(1; \beta, a) + \Gamma(1-a) \Gamma(1; a, \beta) = 2^{1-a/2} \Gamma(1-a) \Gamma(1-\beta)$$

(See [1]). This gives us in particular, the first of the two formulae (6) of ([15], p. 230) with $0 < \text{Re}(a + \beta) < 1$. It seems likely from [16], that there are quite a few of pairs F_1, F_2 which satisfy the conditions of Theorem 3 and for which therefore a Poisson formula holds.

6. A p -adic analogue of the W -operator

Let $\mathcal{F} = \mathcal{F}(\mathbf{Q}_p^\times)$ be now the space of locally constant complex-valued functions F on $\mathbf{Q}_p^\times = \mathbf{Q}_p \setminus \{0\}$ with $F(x) = 0$ for all x with $|x|_p$ sufficiently large and $F(x) = a\mu_1(x) |x|_p^{\frac{1}{2}} + b\mu_2(x) |x|_p^{\frac{1}{2}}$ for all x with $|x|_p$ sufficiently small, where μ_1, μ_2 are quasicharacters of $\mathbf{Q}_p^\times$, $|x|_p$ is a 'normalised' valuation of $\mathbf{Q}_p$ and a, b are complex constants. Such spaces $\mathcal{F}$ occur as "Kirillov models" $\mathcal{K}(\pi)$ for irreducible admissible representations π_p of $GL_2(\mathbf{Q}_p)$; associated with π_p , we

have the L -function $L(s, \pi_p) = \{(1 - \mu_1(p)p^{-(s-\frac{1}{2})})(1 - \mu_2(p)p^{-(s-\frac{1}{2})})\}^{-1}$ where $s \in \mathbb{C}$ and more generally,

$$L(s, \chi, \pi_p) = \{(1 - (\mu_1 \chi^{-1})(p)p^{\frac{1}{2}-s})(1 - (\mu_2 \chi^{-1})(p)p^{\frac{1}{2}-s})\}^{-1},$$

for any character χ of $\mathbb{Z}_p^\times = \mathbb{Z}_p \setminus p\mathbb{Z}_p$. If $(MF_\chi)(s) = \int_{\mathbb{Q}_p^\times} F(x) \chi(x) |x|_p^s d^\times x$

for any F in $\mathcal{F}$, then $(MF_\chi)(s)/L(s, \chi, \pi_p)$ is entire in s and in particular, for the identity character χ_0 , $(MF_{\chi_0})(s)$ is in the space $\mathcal{Z}$ defined by Igusa ([10] chapter I, § 5.2). There exists F_0 in $\mathcal{F}$ such that $(M(F_0)\chi)(s) = L(s, \chi, \pi_p)$ (see [7], § 1.14–1.16) and for $\chi = \chi_0$, F_0 is given by

$$F_0(x) = \begin{cases} 0 & \text{if } x \notin p^{-1}\mathbb{Z}_p \\ |x|_p^{\frac{1}{2}} \sum_{i+j=v_p(x)} \mu_1(p^i) \mu_2(p^j) & \text{if } x \in p^{-1}\mathbb{Z}_p \text{ and } x \neq 0, \end{cases}$$

where $v_p(x)$ is defined by $|x|_p = |p|_p^{v_p(x)}$ for $x \neq 0$. This relation can also be proved by using Theorem 5.3 of Igusa ([10], chapter I) taking $\Lambda = \{s_1 + \frac{1}{2}, s_2 + \frac{1}{2}\}$ where $\mu_i(x) = |x|_p^{s_i}$, $i = 1, 2$.

Corresponding to F in $\mathcal{F}$, let $\mathcal{W} = \mathcal{W}_F$ be the Whittaker function on $GL_2(\mathbb{Q}_p)$; then if

$$L_{\mathcal{W}}(g, \chi, s) = \int_{\mathbb{Q}_p^\times} \mathcal{W} \left(\begin{pmatrix} x & 0 \\ 0 & 1 \end{pmatrix} g \right) \chi^{-1}(x) |x|_p^{2s-1} d^\times x,$$

or $g \in GL_2(\mathbb{Q}_p)$, we have the functional equation

$$\begin{aligned} L_{\mathcal{W}} \left(\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \chi^{-1}, 1-s \right) &= \epsilon(s, \chi, \pi_p) \frac{L(1-s, \chi^{-1}, \pi_p)}{L(s, \chi, \pi_p)} \times \\ &\times L_{\mathcal{W}} \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \chi, s \right) \end{aligned} \quad (27)$$

where $\epsilon(s, \chi, \pi_p)$ is of the form cp^{-vs} for suitable constants c, v . By using the inverse M^{-1} of the Mellin transform M in Theorem 5.3 (Chapter I) of Igusa [10], the W -operator in the space $\mathcal{F}(\mathbb{Q}_p^\times)$ can be defined by rewriting the functional equation above as

$$\frac{(M(WF)\chi^{-1})(1-s)}{L(1-s, \chi^{-1}, \pi_p)} = \epsilon(s, \chi, \pi_p) \frac{(MF_\chi)(s)}{L(s, \chi, \pi_p)},$$

(see also § 1.3, [7]). The function $\epsilon(s, \chi, \pi_p)$ satisfies of course the condition $\epsilon(s, \chi, \pi_p) \epsilon(1-s, \chi^{-1}, \pi_p) = 1$ (see [7]) and therefore the W -operator is of order 2

7. An adelic version of the Poisson formula

Let $F_\infty \in \mathcal{F}(\mathbb{R}_+^\times)$ such that for its Mellin transform $(MF_\infty)(s)$, the quotient $(MF_\infty)(s)/L(s, \pi_\infty)$ is an entire function of s , where [5]

$$L(s, \pi_\infty) = \begin{cases} (2\pi)^{-s-(+1)/2} \Gamma(s + (p+1)/2), & p \geq 0 \text{ in } \mathbb{Z} \\ \pi^{-s-v} \Gamma((s+v)/2) \Gamma((s-v)/2), & v \in \mathbb{C}. \end{cases}$$

Define WF_∞ by

$$\frac{(M(WF_\infty))(s)}{L(s, \pi_\infty)} = \frac{(MF_\infty)(1-s)}{L(1-s, \pi_\infty)}.$$

Let $\mathbf{Q}_A$ be the ring of $\mathbf{Q}$ -adeles and $\mathbf{Q}_A^\times$ the group of $\mathbf{Q}$ -ideles; denote elements x of $\mathbf{Q}_A^\times$ by $(x_\infty, \dots, x_p, \dots)$ with $x_\infty \in \mathbb{R} \setminus \{0\}$ and $x_p \in \mathbf{Q}_p \setminus \{0\}$ and write $\mathbf{Q}^\times$ for $\mathbf{Q} \setminus \{0\}$. Let π_p be irreducible unitary representations of $GL_2(\mathbf{Q}_p)$ for primes p such that the tensor product $\pi = \pi_\infty \otimes_p \pi_p$ gives an irreducible unitary representation of $GL_2(\mathbf{Q}_A)$ and further

$$\prod_p L(s, \pi_p) = \sum_{n \in \mathbb{Z} \setminus \{0\}} a_n |n|^{-s}$$

is a Dirichlet series converging in a right s -half-plane and for every $\chi \in \widehat{\prod_p \mathbf{Z}_p^\times}$,

$$L(s, \chi, \pi) := L(s, \pi_\infty) \prod_p L(s, \chi, \pi_p) = \prod_p \epsilon(s, \chi, \pi_p) L(1-s, \chi^{-1}, \pi). \quad (27)$$

In particular, the functional equation implies that $\prod_p L(s, \pi_p)$ is at most of order

$|s|^r$ for some constant $r = r(B)$ in vertical strips B . On the other hand, $MF_\infty(s)$ and $L(s, \pi_\infty)$ are rapidly decreasing at infinity in vertical strips. Let W_p^0 be the Whittaker function [7] on $GL_2(\mathbf{Q}_p)$ whose Mellin transform (with respect to $\mathbf{Q}_p^\times$) is precisely $L(s, \pi_p)$ for every prime p . Define for $x = (x_\infty, \dots, x_p, \dots) = (x_\infty, x_f)$ in $\mathbf{Q}_A^\times$, the function F on $\mathbf{Q}_A^\times$ by

$$F(x) = F_\infty(|x_\infty|) \prod_p W_p^0 \left(\begin{pmatrix} x_p & 0 \\ 0 & 1 \end{pmatrix} \right) = F_\infty(|x_\infty|) F_f(x_f)$$

and the function WF on $\mathbf{Q}_A^\times$ by

$$(WF)(x) = WF_\infty(|x_\infty|) \prod_p W_p^0 \left(\begin{pmatrix} x_p & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \right).$$

From the properties of F , the series

$$\sum_{\xi \in \mathbf{Q}^\times} F(x\xi)$$

converges absolutely to a function $\varphi_F(x)$ and further $\varphi_F(x)$ is rapidly decreasing as $|x_\infty|$ tends to ∞ or 0 (see [5]). Thus for every character χ on $\mathbf{Q}^\times \setminus \mathbf{Q}_A^\times$ and for $\text{Re}(s)$ sufficiently large,

$$L_F \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \chi, s \right) = \int_{\mathbf{Q}^\times \setminus \mathbf{Q}_A^\times} \varphi_F(x) \chi^{-1}(x) |x|^{2s-1} d^\times x$$

$$= \int_{\mathbf{Q}_A^\times} F(x) \chi^{-1}(x) |x|^{2s-1} d^\times x,$$

is holomorphic in s and admits of a meromorphic continuation to the whole plane; for $\chi \neq \chi_0$ (the identity) it represents an entire function of s . For large enough $\text{Re}(s)$, it is just

$$(MF_\infty)(s) \prod_p L(s, \chi, \pi_p) = (MF_\infty)(s) \sum_{n \neq 0} \chi(n) a_n |n|^{-s};$$

in vertical strips the first factor is rapidly decreasing at infinity (by Theorem 1) while the second factor is at most of order $|s|^a$ for some a (in view of the functional equation and Stirling's formula). Thus $L_F \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \chi, s \right)$ is rapidly decreasing at infinity in vertical strips.

From the local functional equations (27) and (27)', we get

$$L_{WF} \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \chi^{-1}, 1-s \right) = L_F \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \chi, s \right). \quad (28)$$

Let us define for $x \in \mathbf{Q}_A^\times$.

$$F'(x) = \frac{1}{2\pi i} \sum_{\chi'} \int_{\text{Re } s = \sigma} L_F \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \chi', s \right) |x|^{1-2s} ds$$

$$\text{and} \quad F''(x) = \frac{1}{2\pi i} \sum_{\chi'} \int_{\text{Re } s = \sigma_1} L_{WF} \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \chi'^{-1}, 1-s \right) |x|^{1-2s} ds$$

where σ and $-\sigma_1$ are sufficiently large and χ' runs over characters of $\prod_p \mathbf{Z}_p^\times$.

Because of the properties of the integrand mentioned above and the functional equation (28), we obtain by the usual argument

$$F'(x) - F''(x) = \text{Sum of the residues of } L_F \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \chi_0, s \right) |x|^{1-2s}$$

(at the poles encountered while shifting the integration from the far right to the far left, i.e. at all the poles).

$$\text{Since} \quad L_F \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \chi_0, s \right) = \frac{MF_\infty(s)}{L(s, \pi_\infty)} \left(L(s, \pi_\infty) \sum_{n \neq 0} a_n |n|^{-s} \right),$$

the poles arise from the function inside the simple brackets; the residue can be computed and seen to be of the form c_0 or $c_1 \mu_1(x_\infty) + c_2 \mu_2(x_\infty)$ with

constants c_0, c_1, c_2 , depending on the nature of π_∞ . The last relation involving $F'(x) - F''(x)$ may thus be viewed as the adelic formulation of the Poisson formula. When no poles are encountered, it reduces to an 'automorphic' relation $F'(x) = F''(x)$.

References

- [1] Bateman H 1953 *Higher transcendental functions* (New York: McGraw-Hill), Vol. 1, Bateman Manuscript Project
- [2] Berndt B C 1969 Identities involving the coefficients of a class of Dirichlet series I. *Trans. Am. Math. Soc.* **137** 345-359
- [3] Bochner S 1951 Some properties of modular relations. *Ann. Math.* **53** 332-363
- [4] Chandrasekharan K and Raghavan Narasimhan 1961 Hecke's functional equation and arithmetical identities. *Ann. Math.* **74** 1-23
- [5] Gelbart S 1975 Automorphic forms on Adele groups. *Ann. Math. Studies* No. 83 (Princeton University Press)
- [6] Jacquet H and Langlands R P 1970 Automorphic forms on $GL(2)$. *Lecture Notes in Mathematics* (New York: Springer-Verlag)
- [7] Godement R 1970 Notes on Jacquet-Langlands' Theory. *Lecture Notes*, Institute for Advanced Study.
- [8] Hecke E 1936 Über die Bestimmung Dirichletscher Reihen durch ihre Funktionalgleichungen. *Math. Ann.* **112** 664-699; *Gesamm. Abhand* 591-626
- [9] Igusa J.-I 1974 Complex powers and asymptotic expansions. I: Functions of certain types. *J. Reine Angew. Math.* **268-269** 110-130
- [10] Igusa J.-I 1978 *Lectures on forms of higher degree*. Tata Institute of Fundamental Research
- [11] Kubota T 1974 On an analogy to the Poisson summation formula for generalised Fourier transformation. *J. Reine Angew. Math.* **268-269** 180-189
- [12] Maass H 1949 Automorphe Funktionen und indefinite quadratische Formen. Sitz.-ber. Heidelberg. *Akad. Wiss.* 1-42
- [13] Maass H 1949 Über eine neue Art von nichtanalytischen automorphen Funktionen und die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen. *Math. Ann.* **121** 141-183
- [14] Maass H 1953 Die Differentialgleichungen in der Theorie der elliptischen Modul-funktionen. *Math. Ann.* **125** 233-263
- [15] Maass H 1964 Lectures on modular functions of one complex variable. Tata Institute of Fundamental Research
- [16] Ranga Rao R 1977 Unitary representations defined by boundary conditions—the case of $sl(2, \mathbb{R})$. *Acta. Math.* **139** 185-216
- [17] Sally P J 1967 Analytic continuation of the irreducible unitary representations of the universal covering group of $SL(2, \mathbb{R})$. *Mem. Am. Math. Soc.* No. 69
- [18] Weil A 1971 Dirichlet series and automorphic forms. *Lecture Notes in Mathematics* No. 189 (Springer-Verlag)
- [19] Yamazaki T 1978 On a generalisation of the fourier transformation. *J. Fac. Sci. Univ. Tokyo* **25** 237-252

School of Mathematics,
Tata Institute of Fundamental Research,
Bombay 400 005

Orthogonal and spin bundles over hyperelliptic curves

By

S. RAMANAN*

1. Introduction

Let $\Sigma x_i^2 = 0$, $\Sigma \lambda_i x_i^2 = 0$ be two quadrics in the projective space of dimension $(2g + 1)$. Assume that λ_i 's are distinct so that these two quadrics determine a *generic* pencil of quadrics. One can associate to this pencil, in a natural way, a hyperelliptic curve X of genus g , namely, the double of the projective line parametrising the pencil with ramification at the points corresponding to the singular quadrics. One can easily check that the variety M_d of d dimensional vector subspaces which are isotropic for both the quadrics is nonsingular. M_d is nonempty if $d \leq g$ and in [1] we gave an interpretation of M_g and M_{g-1} as the Jacobian and the moduli space of vector bundles of rank 2 and fixed determinant of odd degree over X respectively. One might naturally ask if one could obtain similar interpretation for M_d for other d as well. In particular, does M_1 = the intersection of two quadrics have an interpretation as a moduli-space of some sort of bundles over X ? This paper answers, among other things, this question. In fact, we give a solution of a more general question, namely, the question of interpreting M_d modulo $\Pi SO(n_i)$, when the λ_i 's are allowed to have multiplicities n_i .

The idea of associating a hyperelliptic curve to a pencil of quadrics is nothing new and goes back, at least to Weil [9], who considered the relationship between the intersection of two quadrics (M_1 in our notation) and the curve X from the point of view of diophantine equations over finite fields. He computed the zeta function of M_1 in terms of that of X and verified the Weil conjectures for this variety. Incidentally he also raised the specific question of the geometric relationship between M_1 and X , which we have investigated here. It must also be mentioned that Gauthier [2] has already noticed the identity of the space M_g and the Jacobian of X , although as far as the present author is aware, no proof has been published till recently [6, 1]. The author is thankful to A. Weil for bringing to his attention these references.

The method of proof is quite simple and is not basically different from that in [1]. However, the greater generality leads to conceptual clarity and we have also made a few technical simplifications. If E is a bundle to which the action of i lifts, the bundle does not nevertheless go down to a bundle on $X/i = \mathbf{P}^1$. The obstruction to this is the lack of descent data at the fixed points of i , namely, the

*Supported in part by NSF grant MCS77-18723.

Weierstrass points of X . In other words, i may not act as identity on the fibres of E at the Weierstrass points. To rectify this, one makes a modification of E at these points to obtain a new bundle E' which does go down. If one keeps track of the additional structures needed to recover E from this bundle on $\mathbf{P}^1$ obtained by descent from E' , one can hope for a classification of the bundles started with.

Another description of the method is as follows. Suppose we wish to classify vector bundles of fixed large degree. Consider pairs (E, η) where E is such a vector bundle and η is a trivialisation of E on a sufficiently large divisor D . Then the restriction $H^0(X, E) \rightarrow H^0(D, E)$ is injective and using the trivialisation of E on D , we associate to E , a point of a suitable Grassmannian. Thus one could even construct the moduli space of vector bundles by passing to the quotient of the Grassmannian in k^{dn} by $\Pi GL(d)$. The special situation in our case then allows us to describe the image precisely.

2. Clifford bundles

If V is a vector space of finite dimension and Q is a non-degenerate quadratic form on V , then the Clifford group $\Gamma(Q)$ is the subgroup of the group $C(Q)^\times$ of units in the Clifford algebra $C(Q)$, defined by

$$\Gamma(Q) = \{x \in C(Q)^\times : xVx^{-1} \subset V\}.$$

Let $\Gamma^+(Q)$ be the group of even elements in $\Gamma(Q)$. The multiplicative group $k^\times$ is clearly contained in $\Gamma^+(Q)$. It is actually the centre of $\Gamma(Q)$ if V is of even dimension. Then the spinor norm gives a surjective homomorphism $Nm : \Gamma(Q) \rightarrow k^\times$ whose kernel is called the reduced Clifford group and denoted $\Gamma_0(Q)$. Let $\Gamma_0^+(Q) = \Gamma_0(Q) \cap \Gamma^+(Q)$. By definition, we have a natural (vector) representation of all these groups in the orthogonal group of Q . The kernel of this representation of Γ is $k^\times$. We first notice

Lemma 2.1. If E is an $O(2n)$ -bundle on a scheme X , then the obstruction to its liftability to a $\Gamma_0(2n)$ -bundle (resp. $\Gamma(2n)$ -bundle) is an element of $H^2(X, \mathbf{Z}/2)$ (resp. $H^2(X, \mathcal{O}_X^\times)$). In particular, any $O(2n)$ bundle on a projective nonsingular curve can be lifted to a $\Gamma(2n)$ -bundle. Similar statements are valid for the odd dimensional case when O is replaced by SO and Γ is replaced by Γ^+ .

Proof. This is a trivial consequence of the exact sequences

$$\begin{array}{ccccccc} 1 & \rightarrow & \mathbf{Z}/2 & \rightarrow & \Gamma_0(2n) & \rightarrow & O(2n) \rightarrow 1 \\ (2.1) & & \downarrow & & \downarrow & & \parallel \\ 1 & \rightarrow & k^\times & \rightarrow & \Gamma(2n) & \rightarrow & O(2n) \rightarrow 1, \end{array}$$

and the induced exact sequence of sheaves

$$\begin{array}{ccccccc} 1 & \rightarrow & \mathbf{Z}/2 & \rightarrow & \Gamma_0(2n) & \rightarrow & O(2n) \rightarrow 1 \\ (2.2) & & \downarrow & & \downarrow & & \parallel \\ 1 & \rightarrow & \mathcal{O}_X^\times & \rightarrow & \Gamma(2n) & \rightarrow & O(2n) \rightarrow 1, \end{array}$$

in view of [7]. In fact, the obstruction to the $\Gamma(2n)$ -lifting is actually in the image of $H^2(X, \mathbb{Z}/2)$ in $H^2(X, \mathcal{O}_X^\times)$. Hence our assertion.

If P is a $\Gamma(n)$ -bundle and ξ is a line bundle, then using the action of $k^\times$ on $\Gamma(n)$, we get a new $\Gamma(n)$ -bundle which we will denote $\xi \circ P$. From the fact that $Nm(\lambda a) = \lambda^2 Nm(A)$ for $\lambda \in k^\times$, $A \in \Gamma(n)$ we obtain

$$(2.3) \quad Nm(\xi \circ P) = \xi^2 \otimes NmP.$$

Lemma 2.4. If P, P' are $\Gamma(2n)$ -bundles on X , giving rise to isomorphic $\mathcal{O}(2n)$ -bundles, then there exists a line bundle ξ such that $P' \cong \xi \circ P$. If E is an $\mathcal{O}(2n)$ -bundle and P any lift of E to a $\Gamma(2n)$ -bundle, then NmP is of even or odd degree according as E can or cannot be lifted to a $\Gamma_0(2n)$ -bundle. Similar results hold for the odd dimensional case with $\mathcal{O}(n)$ replaced by $SO(n)$.

Proof. The first assertion follows from the sequence 2.2 and [7]. From this it is clear that (a) the parity of $\deg(NmP)$ is independent of the lift P , and (b) if $\deg NmP$ is even we might, in view of 2.3, as well assume that NmP is trivial by replacing P by a suitable $\xi \circ P$. If NmP is trivial, then P is a $\Gamma_0(2n)$ -bundle as is seen from the sequence

$$1 \rightarrow \Gamma_0(2n) \rightarrow \Gamma(2n) \xrightarrow{Nm} \mathcal{O}^\times \rightarrow 1.$$

Definition 2.5. An $\mathcal{O}(2n)$ -bundle on X is said to be of *even* or *odd type* according as it is or is not liftable to a $\Gamma_0(2n)$ -bundle.

Remark 2.6. An $SO(n)$ -structure on an $\mathcal{O}(n)$ -bundle E with trivial determinant is given by a trivialisation of $\tilde{\Delta}E$, such that its square is the trivialisation of $(\tilde{\Delta}E)^2$ given by the 'discriminant', namely, the lift $\tilde{\Delta}E \rightarrow \tilde{\Delta}E^*$ of the given quadratic form $E \rightarrow E^*$. Accordingly, even when $\tilde{\Delta}E = \det E$ is only an element of order 2, we could still talk of a *special orthogonal structure*, when we mean an isomorphism $\det E \rightarrow \alpha$ such that its square $(\det E)^2 \rightarrow \alpha^2 \simeq 1$, is the trivialisation given by the discriminant. Also, if P is a $\Gamma(n)$ -bundle, we talk of a *special $\Gamma(n)$ -structure* when the associated $\mathcal{O}(n)$ -bundle E comes provided with a special orthogonal structure. If $\det E$ is trivial, this simply means that P is a $\Gamma(n)^+$ -bundle.

3. Orthogonal bundles over hyperelliptic curves

We will make the convention that when we deal with orthogonal bundles of odd rank, we assume that it has trivial determinant. Generally, we will deal with the case of even rank, the odd case being similar.

Let E be an orthogonal bundle over a hyperelliptic curve with a lift $E \approx i^*E$ of the involution i . If P is a lift of E to a Γ -bundle, then we must have $P \approx \alpha \circ i^*P$ for some line bundle α by Lemma 2.4. Taking norms on both sides and using 2.3, we get $\alpha^2 \otimes NmP \cong i^*NmP$ and hence $(\alpha \otimes NmP)^2 \approx h^{\deg NmP}$. In particular, α is of degree 0. Let us then assume given an isomorphism

$$P \approx \beta \circ (NmP)^{-1} \circ i^*P, \text{ with } \beta^2 = h^{\deg NmP}.$$

It is clear that β is essentially unique, at least if NmP is fixed. We now recall

Lemma 3.1. There is a one-one correspondence between line-bundles η of fixed even (respectively odd) degree such that $\eta^2 \approx h^{\deg \eta}$, and the set of partitions $W = S \cup T$ of the set of Weierstrass points into subsets of even (resp. odd) cardinality.

Proof. See ([1] lemma 2.1).

If an isomorphism $i^*E \rightarrow E$ is given, then at every fixed point of i , namely, a Weierstrass point, we obtain an involution of E_w preserving the quadratic form q_w on E_w . Let E_w^+ , E_w^- be the eigenspaces corresponding to the eigenvalues ± 1 . Then E_w^+ , E_w^- are clearly orthogonal for q_w . In particular, q_w is nondegenerate on both these spaces.

We now seek to prove that for such an $O(2n)$ -bundle E , the line bundle β with $i^*P \approx \beta^{-1} \otimes NmP \circ P$ is the unique line bundle with $i^*\beta \approx \beta$ whose associated partition of W (lemma 3.1) is determined by the integers $r_w = \dim E_w^+$. Before we do this, let us normalise the isomorphism $\varphi : P \rightarrow \alpha \circ i^*P$ given above. We will of course require that this should induce the given isomorphism $E \rightarrow i^*E$. This requirement fixes φ up to an automorphism of the $\Gamma(2n)$ -bundle P inducing the identity map on E . This means that φ is determined upto a nonzero scalar factor. We now wish to further specify φ by requiring that it be an involution in the following sense. Firstly, we have an isomorphism $Nm\varphi : NmP \rightarrow \alpha^2 \circ i^*NmP$. Writing $\alpha = \beta \otimes NmP$, this gives an isomorphism $\beta^{-2} \rightarrow NmP \otimes i^*NmP$. On the other hand $i^*\varphi$ gives an isomorphism $i^*P \rightarrow i^*\beta \otimes i^*NmP \circ P$ and hence $\beta \otimes NmP i^*P \approx \beta \otimes i^*\beta \otimes NmP \otimes i^*NmP \circ P$ and using $Nm\varphi$ also $\beta \otimes NmP \circ i^*P \approx \beta^{-1} \otimes i^*\beta \circ P$. Composing this with φ we get an isomorphism $P \rightarrow \beta^{-1} \otimes i^*\beta \circ P$ which induces identity on E . We require that the involutive lift $\beta \approx i^*\beta$ is so chosen that the above composite is the Identity. We shall call this the *normalised isomorphism*.

Let then $\varphi : P \rightarrow \beta \circ NmP \circ i^*P$ be a normalised lift of the given involutive isomorphism $E \rightarrow i^*E$.

If we trivialise β_w at $w \in W$, we get a bijection $\varphi_w : P_w \rightarrow P_w$ which satisfies $\varphi_w(\xi s) = \varphi_w(s)(Nms) \cdot s$. Since P_w is a set on which $\Gamma(2n)$ acts simply transitively we see that φ_w gives rise to an element g_w of $\Gamma(2n)$ which is determined upto the equivalence relation

$$g \sim g' \text{ if there exists } s \in \Gamma(2n) \text{ with } sgs^{-1} \cdot Nms = g'.$$

Moreover g_w gives rise to a reflection in an r_w -dimensional subspace in the vector representation, and satisfies in view of our normalisation $g_w^2 = \epsilon_w Nm g_w$, where $\epsilon_w = \pm 1$ according as the lift of i to β_w is ± 1 at w . Thus we have the following situation. E_w^+ is an r_w -dimensional subspace of E_w and $g_w \in \Gamma(2n, q_w)$ which represents the reflection in E_w^+ satisfies $g_w^2 = \epsilon_w Nm g_w$.

Lemma 3.2. If $g \in \Gamma(2n)$ represents the reflection in a nondegenerate subspace of dimension r , then $g^2 = Nm g$ if $r \equiv 1, 2n$ and $g^2 = -Nm g$ if $r \equiv 3, 2n - 2 \pmod{4}$.

Proof. Let $e_1, \dots, e_{2n}$ be an orthonormal basis for q_w such that $e_1, \dots, e_r$ is a basis for E_w^+ . All the elements e_i belong to $\Gamma(2n)$ and in particular the element $e_1, \dots, e_r$ as well. The induced transformation on E_w is $x \mapsto (e_1, \dots, e_r) x (e_1 \dots e_r)^{-1} =$

$e_1 \dots e_r x e_r \dots e_1$ since $e_r^2 = q_w(e_r) = 1$ by choice. If x is e_i , $i > r$, then $x \mapsto (-1)^r x$ and if x is e_i , $i \leq r$, then $x \mapsto (-1)^{r-1} x$. Thus $g = e_1 \dots e_r$ or $e_{r+1} \dots e_{2n} \in \Gamma(2n)$ represents reflection in E_w^+ according as r is odd or even. Now $(e_1 \dots e_r)^2 = (-1)^{r(r-1)/2}$, while $Nm(e_1 \dots e_r) = 1$. If r is odd, $g^2 = (-1)^{r(r-1)/2} Nm g$. This proves the assertion.

We have now proved

Proposition 3.3. Let E be an orthogonal vector bundle of dimension $2n$ to which the action of i lifts. If P is a Clifford bundle lifting E , then $P \approx \beta \circ (NmP) \circ i^*P$, where β is the unique line bundle of degree $= -\text{degree } NmP$ with $i^*\beta \approx \beta$ associated to the partition of W given by the subset

$$S = \{w \in W : r_w \equiv 1, 2n \bmod 4\},$$

$$T = \{w \in W : r_w \equiv 3, 2n - 2 \bmod 4\},$$

where r_w is the dimension of the fixed subspace of E_w under the action of i . In particular, if E is of even (resp. odd) type, then $\# S$ is even (resp. odd).

Remark 3.4. We call attention to a subtlety regarding special orthogonal structures. Suppose α is a line bundle with an action of i on it. Let E be an orthogonal bundle of odd type with $i^*E \approx E$. Then i^*E also comes with a special orthogonal structure whenever E is provided with one. However, we claim that the isomorphism $\phi : i^*E \rightarrow E$ cannot preserve this special orthogonal structure. For, if it does, we would have a commutative diagram

$$\begin{array}{ccc}
 i^*\det E & \xrightarrow{\det \phi} & \det E \\
 i^*\eta \downarrow & & \downarrow \eta \\
 i^*\alpha & \xrightarrow{i} & \alpha
 \end{array}$$

Restricting to a fibre at $w \in W$, we get a commutative diagram

$$\begin{array}{ccc}
 \det E_w & \xrightarrow{\det \phi_w} & \det E_w \\
 \eta_w \downarrow & & \downarrow \eta_w \\
 \alpha_w & \xrightarrow{\epsilon_w} & \alpha_w
 \end{array}$$

Now, $\det \phi_w = \pm 1$ according as $\dim E_w^-$ is even or odd and since E is of odd type, $\{w : \det \phi_w = 1\}$ is of odd cardinality, while $\{w : \epsilon_w = 1\}$ is of even cardinality (lemma 3.1).

4. Stability of orthogonal bundles

Definition 4.1. Let E be an orthogonal bundle. We say that E is stable (resp. semistable), if every proper isotropic sub-bundle F of E has degree < 0 (resp. ≤ 0).

Proposition 4.2. E is semistable as an orthogonal bundle if and only if it is semistable as a vector bundle.

Proof. It is obvious that semistability (resp. stability) as a vector bundle implies semistability (resp. stability) as an orthogonal bundle. As for the converse, let F be a sub-bundle of E . Consider the sheaf $N = F \cap F^\perp$ where $F^\perp$ is obtained from F by orthocomplementation. Now N may not be a vector sub-bundle of E , but let N' be the sub-bundle generated by it. Then we have the exact sequence

$$0 \rightarrow N' \rightarrow F \oplus F^\perp \rightarrow M' \rightarrow 0,$$

where M' is the sub-bundle of E generated by $F + F^\perp$. Obviously, $M' = (N')^\perp$. Thus $\deg(F \oplus F^\perp) = \deg N' + \deg (N')^\perp = 2 \deg N'$, since we have the exact sequence $0 \rightarrow (N')^\perp \rightarrow E \rightarrow (N')^* \rightarrow 0$. Hence $\deg F = \deg N' \leq 0$, since N' is obviously isotropic.

Remarks 4.3.

(i) It is easy to see that our definition of stability and semi-stability coincides with the general notion introduced by Ramanathan [5], for principal bundles with reductive structure group.

(ii) E could be stable as an orthogonal bundle and yet be nonstable as a vector bundle. The reason is, with the notation of proposition 4.2, that the bundle $F + F^\perp$ could be the whole of E , and $N' = 0$. When this happens, the bundle E is a direct sum of two orthogonal bundles. Actually, this phenomenon does happen as the following example shows.

Example 4.4. Let E be a vector bundle over X of rank 2. Then $\text{ad } E$, the bundle of endomorphisms of trace 0 is a special orthogonal bundle, the Killing form giving the orthogonal structure. It is easy to see, using the one-one correspondence between isotropic sub-bundles of $\text{ad } E$ and sub-bundles of E , that E is stable if and only if $\text{ad } E$ is stable as an orthogonal bundle. On the other hand, if $E \approx \alpha \otimes E$ for some line bundle with $\alpha^2 = 1$, then $\text{ad } E$ contains α as a proper nondegenerate sub-bundle. Indeed, one can even construct bundles (see [4]) with $E \approx \alpha \otimes E$ and $E \approx \beta \otimes E$ for two distinct elements α, β with $\alpha^2 = \beta^2 = 1$, so that $\text{ad } E$ becomes a direct sum $\alpha \oplus \beta \oplus \alpha\beta$.

Regarding this, we have

Proposition 4.5. (i) E is stable as an orthogonal bundle if and only if E is an orthogonal direct sum of sub-bundles E_i , which are mutually nonisomorphic with each E_i stable as a vector bundle.

(ii) If $\text{Aut } E$ is the bundle of orthogonal automorphisms of a stable bundle E , then E is stable as a vector bundle if and only if $\Gamma(\text{Aut } E) = \pm 1$.

Proof. (1) We have proved above that if E is not stable as a vector bundle, then E is the orthogonal direct sum of sub-bundles E_1, E_2 of E . Now clearly E_1, E_2 are stable orthogonal so that by induction on rank E , we see that $E \approx \Sigma E_i$, where E_i are stable as vector bundles. If such a direct sum has repeated summands, say $E_1 \simeq E_2$, then the imbedding $X \hookrightarrow (X, iX)$ of E_1 in $E_1 \oplus E_1$ gives an isotropic sub-bundle of degree zero, contradicting the stability of E . Thus all the E_i 's are mutually nonisomorphic. Conversely, if $E = \Sigma E_i$, E_i stable as vector bundles and $E_i \not\sim E_j$, then any sub-bundle $\neq 0$ of degree 0 must be a sum of some of the E_i and hence cannot be isotropic.

(2) If E is stable as a vector bundle, then $\Gamma(\text{Aut } E) \subset k^*$ and hence must be ± 1 . The converse is a consequence of (1).

Proposition 4.6. If E is an orthogonal bundle over a curve with an involution i which lifts to E , then E is semistable as an orthogonal bundle if and only if for every i -invariant, isotropic sub-bundle F of E , we have $\deg F \leq 0$.

Proof. If F is any sub-bundle of E , then as in proposition 4.2, we have the sequence

$$0 \rightarrow N' \rightarrow F \oplus i^*F \rightarrow M' \rightarrow 0,$$

where N' , M' are sub-bundles generated by $F \cap i^*F$ and $F + i^*F$ respectively. Now $\deg N' \leq 0$ by assumption, while the fact that $\deg M' \leq 0$, follows from proposition 4.2. Since $\deg F = \deg i^*F$, this proves our claim.

Let X be a hyperelliptic curve with rational involution i . Let E be a vector bundle on X with a lift of i . For every $w \in W$, the set of Weierstrass points of X , denote by r_w the dimension of the space of fixed points of i on the fibre E_w . Consider the exact sequence

$$(4.7) \quad 0 \rightarrow E' \rightarrow E \rightarrow \sum_{w \in W} E_w^- \otimes \mathcal{O}_w \rightarrow 0,$$

where E_w^- is the eigenspace for i_w corresponding to the eigenvalue -1 . The maps $E \rightarrow E_w^- \otimes \mathcal{O}_w$ are given by the natural projections $E_w \rightarrow E_w^-$. The kernel E' is clearly locally free. Moreover, the involution i acts on E' .

Lemma 4.8. For every $w \in W$, the involution i acts as Identity on E'_w .

Proof. We have only to show that i acts as Identity on the kernel $E'_w \rightarrow E_w$ since it does so on the image which is E_w^+ . Now this kernel is $\text{Tor}_1(E_w^- \otimes \mathcal{O}_w, \mathcal{O}_w) = E_w^- \otimes \text{Tor}_1(\mathcal{O}_w, \mathcal{O}_w)$. From the resolution

$$0 \rightarrow L_w^{-1} \rightarrow \mathcal{O}_w \rightarrow \mathcal{O}_w \rightarrow 0,$$

we conclude that $\text{Tor}_1(\mathcal{O}_w, \mathcal{O}_w)$ is canonically isomorphic to the fibre $(L_w^{-1})_w = K_w$, the space of differentials at w . Since i acts as -1 Identity on K_w , it follows that i acts as Identity on $E'_w = E_w^- \otimes K_w$.

From lemma 4.8, the bundle E' descends to a vector bundle $\tilde{E}$ on $X/i \approx \mathbf{P}^1$. Let us now make the assumption on E :

(A) E satisfies $H^0(X, E \otimes h^{-\sigma}) = 0$ and $H^1(X, E) = 0$. Then we have $H^0(X, E' \otimes h^{-\sigma}) \subset H^0(X, E \otimes h^{-\sigma}) = 0$ by (A) and hence in particular $H^0(\mathbf{P}^1, \tilde{E} \otimes h^{-\sigma}) = 0$. Consequently, $\tilde{E} \otimes h^{-\sigma}$ is a direct sum of line bundles, each of which is of degree < 0 . This implies that $\tilde{E} \otimes h^{-(\sigma-1)}$ is a direct sum of line bundles of degree ≤ 0 and hence contained in a trivial bundle. Cononically speaking, we have $\tilde{E} \otimes h^{-(\sigma-1)} \subset H^1(\tilde{E} \otimes h^{-(\sigma+1)}) \mathbf{P}^1$. On the other hand, we have the isomorphism $H^1(\mathbf{P}^1, \tilde{E} \otimes h^{-(\sigma+1)}) \simeq H^1(X, E)^*$ in view of

Lemma 4.9. For any vector bundle V on $\mathbf{P}^1$, we have the isomorphism $H^1(\mathbf{P}^1, V) \approx H^1(X, \pi^*V \otimes h^{-(\sigma+1)})^*$,

Proof. The left side is dual to $H^0(\mathbf{P}^1, V^* \otimes h^{-2}) \approx H^0(X, \pi^* V^* \otimes h^{-2})^\#$ while the right side is dual to $H^0(X, \pi^* V^* \otimes h^{-(\sigma+1)} \otimes K_X)^\dagger$. But the isomorphism $h^{-(\sigma+1)} \otimes K_X \approx h^{-2}$ commutes with i only upto sign. This proves the assertion.

Thus $\tilde{E} \otimes h^{-(\sigma+1)} \subset H^1(X, E')_{\mathbf{P}^1}$ canonically. From the exact sequence 4.7, we obtain the exact sequence

$$0 \rightarrow H^0(X, E)^\dagger \rightarrow \Sigma E_w^- \rightarrow H^1(X, E')^\dagger \rightarrow 0.$$

To check the injectivity $H^0(X, E)^\dagger \rightarrow \Sigma E_w^-$, we note that its kernel is contained in the kernel $H^0(X, E) \rightarrow \Sigma E_w$ which is $H^0(X, E \otimes L_w^{-1}) = H^0(X, E \otimes h^{-(\sigma+1)}) \subset H^0(E \otimes h^{-\sigma}) = 0$ by assumption (A). As for the surjectivity $\Sigma E_w^- \rightarrow H^1(X, E')^\dagger$, this follows from the fact that $H^1(X, E)^\dagger \subset H^1(X, E) = 0$ by (A).

Let us return to orthogonal bundles. If F is a semistable orthogonal bundle, then $E = F \otimes \alpha$ for some i -invariant line bundle α of degree $j = 2g - 1$ satisfies (A). Thus we can associate to such a bundle the subspace $H^0(X, E)^\dagger$ of ΣE_w^- . The orthogonal structure on F gives rise to nondegenerate quadratic forms on F_w^+ and F_w^- (see § 3), and hence an α_w^2 -values quadratic form on E_w^- . This fits in the following diagram

$$(4.10) \quad \begin{array}{ccc} H^0(X, E)^\dagger & \xrightarrow{\quad\quad\quad} & \Sigma E_w^- \\ \downarrow q & & \downarrow q_w \\ H^0(X, \alpha^2)^\# \approx H^0(\mathbf{P}^1, h^j) & \xrightarrow{\quad\quad\quad} & \Sigma \alpha_w^2 = h_w^j \end{array}$$

Now consider the exact sequence (on $\mathbf{P}^1$)

$$0 \rightarrow h^{j-(2\sigma+2)} \rightarrow h^j \rightarrow h^j | \mathcal{O}_w \rightarrow 0.$$

Then we get a linear map $\Sigma h_w^j \rightarrow H^1(\mathbf{P}^1, h^{j-(2\sigma+2)})$. Clearly, $\dim H^1(\mathbf{P}^1, h^{j-(2\sigma+2)}) = 1$ or 2 according as $j = 2g$ or $2g - 1$.

In either case, we have

Proposition 4.11. The quadratic form on E_w^- with values in h_w^j gives rise to a form on ΣE_w^- or a quadratic map with values in a 2-dimensional space according as $j = 2g$ or $2g - 1$. In any case, the subspace $H^0(X, E) \subset \Sigma E_w^-$ is totally isotropic for the quadratic maps involved.

Proof. This is a simple consequence of the commutativity of 4.10, and the exactness of

$$H^0(\mathbf{P}^1, h^j) \rightarrow \Sigma h_w^j \rightarrow H^1(\mathbf{P}^1, h^{j-(2\sigma+2)}).$$

Let us further analyse the situation. If $j = 2g$, a simple computation using ([1], proposition 2.2) shows that $H^0(X, E)^\dagger$ has dimension $\frac{1}{2}(\Sigma \dim E_w^-)$. In other words, it is a maximal isotropic space for the quadratic form. If $j = 2g - 1$, $H^0(X, E)^\dagger$ has dimension $= \frac{1}{2}(\Sigma \dim E_w^- - rkE)$. For any $t \in \mathbf{P}^1 - W$, we get a linear map $H^1(\mathbf{P}^1, h^{-3}) \rightarrow H^1(\mathbf{P}^1, h^{-2})$ and hence a nondegenerate quadratic form on ΣE_w^- . It is easy to determine the orthocomplement of $H^0(X, E)^\dagger$ for this form. Indeed, it is $H^0(X, E \otimes h)^\dagger$ where the latter space is imbedded by evaluation in $\Sigma E_w^- \otimes h_w$, and h_w is trivialised by means of the section of h vanishing at the point of X lying over t . To see this, we have only to show that the bilinear

product of an element in $H^0(X, E)$ and an element in $H^0(X, E \otimes h)^\dagger$ goes to zero in $H^1(\mathbf{P}^1, h^{-2})$. From the diagram

$$\begin{array}{ccccc} H^0(X, h^{2\sigma-1}) & \rightarrow & \Sigma h_w^{2\sigma-1} & \rightarrow & H^1(X, h^{-3}) \\ \downarrow & & \downarrow & & \downarrow \\ H^0(X, h^{2\sigma}) & \rightarrow & \Sigma h_w^{2\sigma} & \rightarrow & H^1(X, h^{-2}) \end{array}$$

this is equivalent to showing that q_w maps it in the image of $H^0(X, h^{2\sigma})$ in $\Sigma h_w^{2\sigma}$. But this is obvious from the diagram

$$\begin{array}{ccc} H^0(X, E)^\dagger \otimes H^0(X, E \otimes h) & \longrightarrow & \Sigma E_w^- \\ \downarrow & & \downarrow \\ H^0(X, a^2 \otimes h)^\dagger = H^0(\mathbf{P}^1, h^{2\sigma}) & \searrow & \Sigma h_w^{2\sigma-1} \\ & & \swarrow \\ & & \Sigma h_w^{2\sigma} \end{array}$$

In other words, we have shown

Proposition 4.12. The ortho complement of $V = H^0(X, E)^\dagger \subset \Sigma E_w^-$ is $H^0(X, E \otimes h)^\dagger$ for the quadratic form at $t \in \mathbf{P}^1$ and $H^0(X, E \otimes h)^\dagger / H^0(X, E)^\dagger$ is canonically the fibre of $\tilde{E} \otimes h^{-(\sigma-1)}$ at t .

In other words, if the quadratic map $\Sigma E_w^- \rightarrow H^1(\mathbf{P}^1, h^{-2})$ is regarded as a bundle with an h -valued quadratic form on the trivial bundle ΣE_w^- over $\mathbf{P}^1$, then $V^\perp / V \approx \tilde{E} \otimes h^{-(\sigma-1)}$.

Proof. We have only to trace the isomorphisms

$$\begin{aligned} H^0(X, E \otimes h)^\dagger / H^0(X, E)^\dagger &\approx \ker \Sigma E_w^- / H^0(X, E)^\dagger \rightarrow \Sigma E_w^- / H^0(X, E \otimes h) \\ &\simeq \ker H^1(X, E') \rightarrow H^1(X, E' \otimes h) \\ &\approx \ker H^1(\mathbf{P}^1, \tilde{E} \otimes h^{-(\sigma+1)}) \rightarrow H^1(\mathbf{P}^1, \tilde{E} \otimes h^{-\sigma}) \end{aligned}$$

this map being given by the section of h vanishing at t . From the exact sequence (on $\mathbf{P}^1 \times \mathbf{P}^1$)

$$\begin{aligned} 0 \rightarrow p_1^*(\tilde{E} \otimes h^{-(\sigma+1)}) &\rightarrow p_1^*(\tilde{E} \otimes h^{-(\sigma+1)}) \otimes L_\Delta \rightarrow p_1^*(\tilde{E} \otimes h^{-(\sigma+1)}) \\ &\otimes L_\Delta \otimes \mathcal{O}_\Delta \rightarrow 0, \end{aligned}$$

where Δ is the diagonal divisor in $\mathbf{P}^1 \times \mathbf{P}^1$. On the other hand $L_\Delta|_\Delta \approx K_x^-$ where K_x is the canonical line bundle. Taking direct images on $\mathbf{P}^1$, we get

$$E \otimes h^{-(\sigma+1)} \otimes h^2 \approx \text{the kernel required.}$$

5. Classification of i -invariant orthogonal bundles

Theorem 1. Let X be a hyperelliptic curve of genus ≥ 2 . Then the space U of semistable orthogonal bundles of rank r with an i -action is isomorphic to the following space. Let $\Sigma \xi_w$ be an orthogonal direct sum of spaces ξ_w with non-degenerate quadratic forms Q_w . Let ΣQ_w and $\Sigma \lambda_w Q_w$ be two quadratic forms, where

λ_w are mutually distinct scalars determined by X . Then the group $\Pi O(Q_w)$ acts on the variety M of subspaces of dimension $\frac{1}{2}(\Sigma \dim \xi_w - r)$ isotropic for both these forms. The quotient (in the sense of geometric invariant theory) is isomorphic to U .

Proof. Let us for convenience reinterpret the pencil of quadrics on $\Sigma \xi_w$. First of all, we will assume the standard quadratic forms Q_w to have values, not in the field, but in the one-dimensional space h_w^{2g-1} . Then the quadratic map $\Sigma \xi_w \rightarrow \Sigma h_w^{2g-1}$ gives, on composition with the boundary homomorphism $\Sigma h_w^{2g-1} \rightarrow H^0(\mathbf{P}^1, h^{-3})$ of the sequence

$$0 \rightarrow h^{-3} \rightarrow h^{2g-1} \rightarrow h^{2g-1} \otimes \mathcal{O}_W \rightarrow 0,$$

a pencil of quadrics as required.

Fix a line bundle a of degree $2g - 1$ with an i -action on it, once for all. Then we have seen in § 4, that we have a natural inclusion $H^0(X, E) \subset \Sigma E_w^-$, where $E = F \otimes a$, F an orthogonal bundle with an i -action satisfying condition (A) in 4.8. Choosing quadratic isomorphisms $E_w^- \rightarrow \xi_w$, we associate to E , a subspace of $\Sigma \xi_w$ of $\dim = \dim H^0(X, E)^{\frac{1}{2}} = \frac{1}{2}(\dim \xi_w - n)$, where $rk E = n$. By proposition 4.11, we see that this subspace is isotropic for the given pencil on $\Sigma \xi_w$. Moreover, the intersection of this space with ξ_w is the kernel of

$$H^0(X, E)^{\frac{1}{2}} \rightarrow \sum_{w' \neq w} E_{w'},$$

and is hence isomorphic to $H^0(X, E \otimes L_{W-w}^{-1})^{\frac{1}{2}} \subset H^0(X, E \otimes L_{W-w}^{-1}) \subset H^0(X, E \otimes h^{-g}) = 0$ by (A). Thus the subspace V associated to F satisfies (A') V is a subspace of $\Sigma \xi_w$ which is isotropic for the pencil of quadrics and $V \cap \xi_w = 0$ for every $w \in W$.

Conversely, if V is a subspace of $\Sigma \xi_w$ satisfying (A'), then the ortho complement of V with respect to any quadric of the pencil is of the same dimension. Indeed this needs to be checked only for quadrics corresponding to points $w \in W \subset \mathbf{P}^1$. The nullity of this quadric is ξ_w and $V \cap \xi_w = 0$ implies that $V^\perp$ is of the complementary dimension containing ξ_w . This proves that $V^\perp/V$ is a vector bundle on $\mathbf{P}^1$ which comes with a subspace of the fibre $(V^\perp/V)_w$, namely, the isomorphic image of ξ_w . Using this, we can construct a bundle E (on X)

$$0 \rightarrow V^\perp/V \otimes h^{g-1} \rightarrow E \rightarrow \mathcal{G} \rightarrow 0,$$

in which $\mathcal{G}$ is a torsion sheaf concentrated on W with length $(\mathcal{G}) = \dim \xi_w$, and the kernel $(V^\perp/V \otimes h^{g-1})_w \rightarrow E_w$ is $\xi_w \otimes h_w^{g-1}$. Then $E = F \otimes a$ with F an i -invariant orthogonal bundle. It is then easy to see that $H^0(E \otimes h^{-g}) = 0$, so that we have

Proposition 5.1. Let a be a line bundle on X of degree $2g - 1$, invariant under i . There is then a bijection between the set of orthogonal bundles F of rank n together with (a) an i -action and (b) orthogonal isomorphisms $\eta_w(F_w \otimes a_w)^- \rightarrow \xi_w$, satisfying $H^1(X, F \otimes a) = 0$, and the space of subspaces V of $\Sigma \xi_w$ of dimension $\frac{1}{2}(\Sigma \dim \xi_w - n)$ which are isotropic for the pencil of quadrics $\Sigma Q_w = 0$, $\Sigma \lambda_w Q_w = 0$, and satisfy $V \cap \xi_w = 0$ for all w .

Clearly, this bijection is compatible with the action of $\Pi O(Q_w)$ on both the sets. The action on (F, η_w) is simply $(F, g_w \circ \eta_w)$ by $g = (g_w)$. On the other hand, clearly,

the group $\Pi O(Q_w)$ acts on $\Sigma \xi_w$ and leaves the pencil invariant and hence also acts on the set of isotropic subspaces in question.

To complete the proof of theorem 1, we have to investigate the correspondence between stable (resp. semistable) bundles and stable (resp. semistable) points under the action of $\Pi O(Q_w)$.

We recall

Definition 5.2. Let G be a reductive group acting linearly on a vector space and hence also on the projective space. If M is a projective variety invariant under G , a point $m \in M$ is said to be semistable (resp. properly stable) for the action of G , if it satisfies the following equivalent conditions.

(1) There exists a G -invariant hypersurface not passing through m (resp. and in addition, the isotropy group at m is finite and the orbit of m in the complement of this hypersurface is closed).

(2) For any one parameter group $t \mapsto \lambda(t)$ of G , consider $\lim_{t \rightarrow 0} \lambda(t)m \in M$. This point is represented by a line in V on which the one parameter group acts as a character $t \mapsto t^{s_m}$, for some $s_m \in \mathbb{Z}$. Then $s_m \geq 0$ (resp. $s_m > 0$).

The equivalence of these two conditions is proved in ([3], Chapter 2).

For the action of $\Pi O(Q_w)$ on the space of isotropic subspaces, we have

Proposition 5.3. A subspace V of $\Sigma \xi_w$ is semistable (resp. properly stable) if and only if for every proper family (N_w) of isotropic subspaces of ξ_w , we have

$$\dim(\Sigma N_w) \cap V + \dim(\Sigma N_w^\perp) \cap V \leq (\text{resp. } <) \dim V.$$

Proof. We will use the equivalent condition (2) in definition 5.2 to prove this. Any one-parameter group $t \mapsto \lambda_t$ of $\Pi O(Q_w)$ can be described as follows. There exist isotropic subspaces N_i of $\Sigma \xi_w$, compatible with this orthogonal decomposition with

$$0 = N_0 \subset N_1 \subset \dots \subset N_r \subset N_r^\perp \subset N_{r-1}^\perp \subset \dots \subset N_1^\perp \subset N_0^\perp = \Sigma \xi_w = \bar{w}.$$

λ_t leaves all N_i invariant and induces the character $t \mapsto t^{a_i}$ on N_i/N_{i-1} , $i = 1, \dots, r$ with $a_1 \leq a_2 \leq \dots \leq a_r \leq 0$. The action on $N_r^\perp/N_r$ is trivial, while on $N_{i-1}^\perp/N_i^\perp$, it acts as $t \mapsto t^{-a_i}$. Now if V is any isotropic subspace of $\Sigma \xi_w$, then $\lim_{t \rightarrow 0} \lambda_t(V)$ is a space on which λ acts as on the associated graded space for the above filtration (N_i) . In other words, we have

$$\begin{aligned} s_V &= \sum_{i=1}^r a_i \dim \frac{N_i \cap V}{N_{i-1} \cap V} + \sum_{i=1}^r (-a_i) \dim \frac{N_{i-1}^\perp \cap V}{N_i^\perp \cap V} \\ &= \sum_{i=1}^r a_i (\dim N_i \cap V - \dim N_{i-1} \cap V) \\ &\quad + \sum_{i=1}^r (-a_i) \left(\dim \text{Im } V \text{ in } \frac{W}{N_i^\perp} - \dim \text{Im } V \text{ in } \frac{W}{N_{i-1}^\perp} \right) \\ &= \sum_{i=1}^r (a_i - a_{i+1}) \left(\dim N_i \cap V - \dim \text{Im } V \text{ in } \frac{W}{N_i^\perp} \right) \end{aligned}$$

where we make the convention that $a_{r+1} = 0$. Since $a_i - a_{i+1}$ are arbitrary negative integers, we must have

$$\begin{aligned} \dim \left(N_i \cap V - \dim \operatorname{Im} V \text{ in } \frac{W}{N_i^\perp} \right) \\ = \dim (N_i \cap V) + \dim (N_i^\perp \cap V) - \dim V > 0, \end{aligned}$$

in order that we may have $s_V < 0$. This proves our assertion.

We will now investigate the equivalence among semistable points which gives the geometric invariant theoretic quotient. According to [3], two semistable points are equivalent if their orbits have a common limit point. We claim

Proposition 5.4. Two semistable subspaces $V_1, V_2 \subset \Sigma \xi_w$ represent the same point in the quotient if and only if there exist isotropic flags $(N_1), (N_2) \subset \Sigma \xi_w$ compatible with the decomposition such that the associated graded spaces V'_1, V'_2 are in the same G -orbit. Moreover, any semistable subspace V is equivalent to a space $\bigoplus V_i$ where each V_i is stable in $\Sigma V_i \cap \xi_w$.

Proof. If V is not stable, then there exists an isotropic sub-bundle $N \subset \Sigma \xi_w$, compatible with the decomposition such that

$$\dim V \cap N + \dim V \cap N^\perp = \dim V.$$

Now using induction and replacing V by $V \cap N^\perp / V \cap N \subset N^\perp / N$, we can prove the last part of the statement. From the analysis of the stability condition, we see that the closure of an orbit of V under a suitable one-parameter group contains the associated graded of V . Finally, it remains to prove that if V_1, V_2 have different stable series, they map on distinct points in the quotient. This is clear, if V_1 is stable since then, the orbit is closed. In general, we use induction on the length of the stable series and semicontinuity of rank in a family of linear transformations.

Lemma 5.5. If a point $V \subset \Sigma \xi_w$ is semistable for the action of $\Pi O(Q_w)$, then $V \cap \xi_w = 0$ for all w .

Proof. Indeed, take $N_{w_0} = V \cap \xi_{w_0}$, and $N_w = 0$ for $w \neq w_0$ in proposition 5.3. Then semistability of V implies that

$$\dim (V \cap \xi_{w_0}) + \dim (V \cap (N_{w_0}^\perp + \sum_{w \neq w_0} \xi_w)) \leq \dim V.$$

Since V is isotropic, any $v = \Sigma v_w \in V$ should satisfy $v_{w_0} \perp N_{w_0}$. Hence $V \subset N_{w_0}^\perp + \sum_{w \neq w_0} \xi_w$. This leads to a contradiction if $N_{w_0} \neq 0$.

In view of lemma 5.5, we claim we have a natural morphism from the open set of semistable points into $\tilde{M}$, where $\tilde{M}$ is the variety over the moduli space M consisting of i -invariant orthogonal bundles E , together with isomorphisms $E_w^- \rightarrow \xi_w$ as in proposition 5.1. To see this, we need only check.

Proposition 5.6. A semistable (resp. stable) point V corresponds to a semistable (resp. stable) bundle.

Proof. In fact, if F is such that $H^1(X, F \otimes \alpha) = 0$, and if $H^0(F \otimes \alpha)^\dagger \subset \Sigma \xi_w$ is a semistable point, then we have to show that F is semistable. Let $L \subset F$ be any i -invariant isotropic sub-bundle. Then $H^0(X, L \otimes \alpha)^\dagger \subset H^0(X, F \otimes \alpha)^\dagger$ is a subspace contained in $H^0(X, F \otimes \alpha)^\dagger \cap (\Sigma(L \otimes \alpha)_w^-)$. Taking $N_w = (L \otimes \alpha)_w^-$ in the definition of stable points (proposition 5.3), we get

$$\dim H^0(X, L \otimes \alpha)^\dagger + \dim H^0(X, L^\perp \otimes \alpha)^\dagger \leq \dim H^0(X, F \otimes \alpha)^\dagger.$$

This implies that

$$\begin{aligned} \dim H^0(X, L \otimes \alpha)^\dagger - \dim H^1(X, L \otimes \alpha)^\dagger + \dim H^0(X, L^\perp \otimes \alpha)^\dagger \\ - \dim H^1(X, L^\perp \otimes \alpha)^\dagger \leq \dim H^0(F \otimes \alpha)^\dagger. \end{aligned}$$

Using ([1], proposition 2.2), we get (setting $\dim(L \otimes \alpha)_w^- = a_w$, $rk L = r$)

$$r + \frac{1}{2}(\deg L + (2g - 1)r - \Sigma(r - a_w))$$

$$+ n - r + \frac{1}{2}(\deg L^\perp + (2g - 1)(n - r) - \Sigma(n - r - r_w + a_w)) \leq \frac{1}{2}\Sigma r_w - n).$$

i.e., $n + \frac{1}{2}(\deg L + \deg L^\perp + (2g - 1)n - (2g + 2)n + \Sigma r_w) \leq \frac{1}{2}(\Sigma r_w a_w - n)$.

This means that $\deg L + \deg L^\perp \leq 0$, proving our assertion.

We have thus obtained a morphism from the open subset of semistable points into the required moduli space. It is obvious that this morphism is $\Pi\mathcal{O}(Q_w)$ -invariant and hence goes down to a morphism of the quotient variety. Moreover, proposition 5.1 implies that this is an isomorphism of the open set in the quotient corresponding to points representing closed orbits in the set of semistable points (namely stable points), onto an open subset. Hence it follows that there is a surjective, birational morphism of this quotient into the moduli variety.

Now since the moduli variety in question is a component of the action of i on the moduli of orthogonal bundles and since the latter is normal [5], it follows that the former is also normal. Finally to prove the theorem we have only to check that the morphism is injective. Any point in the quotient can be represented by ΣV_i where V_i is stable in $\Sigma V_i \cap \xi_w$. Thus ΣV_i is mapped on the bundle $\Sigma \varphi_i(V_i)$, where φ_i are similar maps defined on $\Sigma V_i \cap \xi_w$. If $\varphi(V)$ is S -equivalent to $\varphi(V')$ then it is necessary that after a permutation of the factors, we must have $\varphi_i(V_i) \sim \varphi_i(V'_i)$. But this implies that V_i and V'_i are in the same orbit, under the group $\Pi\mathcal{O}(V_i \cap \xi_w)$. This proves that φ is injective and hence that it is an isomorphism by Zariski's main theorem.

6. Spin structures and low dimensional cases

If E is an orthogonal bundle of odd type, then any lift P of E to a Γ -bundle has norm of odd degree. Now the lift of i -action to E gives rise to a line bundle β as in §3, determined by the behaviour of i at w . For each such lift P of E with fixed norm, we obtain a lift to $\Gamma(q_w)$ of a reflection in the subspace E_w^+ with respect to the quadratic form q_w induced from the quadratic structure on E . As we have seen in proposition 3.3, this induces an orientation on each E_w^- , and hence we have analogously the

Theorem 2. The moduli of Γ^+ -bundles P of fixed norm of odd degree with isomorphisms $i^*P \rightarrow \alpha \circ P$ of Γ -bundles is isomorphic to the quotient of the variety of isotropic spaces for the pencil

$$\Sigma Q_w = 0, \Sigma \lambda_w Q_w = 0,$$

where Q_w is a nondegenerate form on a space ξ_w , by the group $\Pi SO(Q_w)$.

Proof. What we have to show is only that the space of i -invariant Γ^+ -bundles are in $(1, 1)$ -correspondence with i -invariant orthogonal bundles modulo the group $\Pi SO(E_w^-)$. Now the group $\Pi SO(E_w^-)/\Pi O(E_w)$ is isomorphic to $(\mathbf{Z}/2)^W$ and acts on the set of Γ^+ -bundles of fixed norm with a fixed associated orthogonal bundle E as follows. An element of $(\mathbf{Z}/2)^W$ can be identified (lemma 3.1) with line bundles whose square is 1 or $h^{\deg N}$. Now if $\alpha^2 = 1$, it acts on a Γ^+ -bundle P by sending P to $\alpha \circ P$. If $\alpha^2 \simeq N$, we send P to $\alpha^{-1} \circ N \circ i^*P$. It is easy to see that this action is simply transitive. Indeed, the subgroup $\{\alpha : \alpha^2 = 1\}$ clearly acts on the set of Γ^+ -bundles with a fixed associated special orthogonal bundle E . Any element in the other coset twists P into another Γ^+ -bundle whose associated special orthogonal bundle is not isomorphic to P (remark 3.4). This proves our assertion. Finally, to check that the morphism from the moduli of Γ^+ -bundles to the variety of isotropic spaces is actually an isomorphism (instead of being just bijective), we have to show that the inverse map is an isomorphism. But we know that the composite of the inverse map with the morphism into the moduli of orthogonal bundles is a morphism. Now this follows from the fact that, given any orthogonal bundle in $T \times X$ and $T \in T$, there exists an étale neighbourhood $U \rightarrow T$ of t such that the induced bundle on $U \times X$ can be lifted to a Γ^+ -bundle.

By using lemma 2.1 and the Kunneth decomposition, we can pass to such a neighbourhood U that the induced bundle on $U \times X$ has the obstruction to Γ_0 -lifting in $H^1(X, \mathbf{Z}/2)$. But its image in $H^2(X, \mathbb{O}^*)$ is zero and hence the obstruction to Γ -liftability of the induced bundle is also zero, since it is in the image of $H^2(X, \mathbb{O}^*) \rightarrow H^2(U \times X, \mathbb{O}^*)$. This proves theorem 2.

We will now make a few remarks on the dimensions of ξ_w , although these are implicit.

Remark 6.1. When we deal with orthogonal bundles F of even rank, then the corresponding integers ($\dim F_w^-$) have the property: $\{w : \dim F_w^- \text{ is odd}\}$ is of even cardinality. Thus we see that F has even or odd rank according as $\Sigma \dim \xi_w$ is even or odd. This also follows from ([1], proposition 2.2) since $\frac{1}{2}((2g-1)n - \Sigma \dim E_w^-)$ is an integer, where $E = F \otimes \alpha$, $\deg \alpha = 2g-1$. By construction, ξ_w is isomorphic to E_w^- .

We shall now consider special cases. The simplest is when all the spaces ξ_w are one-dimensional or 0. In that case, there are no proper isotropic subspaces contained in ξ_w . Moreover, the group $\Pi O(Q_w)$ is a finite group. If we deal with $\Gamma(n)$ -structures in such a case, then the moduli space is actually isomorphic to the variety of isotropic spaces. In other words, we have

Theorem 3. The moduli space of i -invariant orthogonal bundles E of rank $2n$ with Γ^+ -structures such that $\dim((E \otimes \alpha)_w^-) = 1$ for all w , where α is an i -invariant line bundle of degree $2g-1$, is isomorphic to the variety of $(g+1-n)$ -dimension subspaces of k^{2g+2} which are isotropic to the quadrics

$$\Sigma X_i^2 = 0, \Sigma \lambda_i X_i^2 = 0.$$

In particular, the intersection of the above two quadrics is itself a moduli space of $\Gamma^+(2n)$ -bundles of the above type.

6.2. *Case of rank 2.* Let us further specialise to the case $\text{rank} = 2$. In this case, i -invariance of orthogonal bundles of determinant 1 is automatic. Thus we get: *The Jacobian of X is isomorphic to the variety of g -dimensional subspaces of k^{2g+2} , isotropic for $\Sigma X_i^2 = 0$ and $\Sigma \lambda_i X_i^2 = 0$.* See [1, 2 and 6].

6.3. *Case of rank 4.* In this case, $\Gamma^+(4)$ is isomorphic to the subgroup of $GL(2) \times GL(2)$ consisting of elements (A, B) with $\det A \cdot \det B = 1$. The homomorphism into $SO(4)$ is given by $(A, B) \mapsto A \otimes B$. On the other hand, $\Gamma(4)$ is the split extension of $\mathbf{Z}/(2)$ by $\Gamma^+(4)$ given by the action $(A, B) \rightarrow (B, A)$, by the non-trivial element of $\mathbf{Z}/(2)$. Thus a $\Gamma^+(4)$ -bundle is essentially a pair of bundles M, N with $\det M \otimes \det N = \text{trivial bundle}$. But a $\Gamma(4)$ -bundle does not distinguish between M and N . Firstly, we have: *The moduli space of vector bundles of rank 2 and fixed determinant of odd degree is isomorphic to the variety of $(g-1)$ -dimensional subspaces of k^{2g+2} , isotropic for $\Sigma X_i^2 = 0$ and $\Sigma \lambda_i X_i^2 = 0$.* The group of even changes of sign in the coordinates correspond to tensorisation by line bundles of order 2.

6.4. *Case of rank 6.* If R is a four-dimensional vector space, ${}^2\Lambda R$ is a six-dimensional space equipped with a ${}^4\Lambda R$ -valued quadratic form. Thus we get a homomorphism of the subgroup $G = \{(\lambda, g) : \lambda^2 = \det g\}$ of $k^* \times GL(R)$ into $SO(6)$ by sending (λ, g) to $\lambda^{-1} \circ {}^2\Lambda g$. It is easy to see that $G = \Gamma^+(6)$. Thus a $\Gamma^+(6)$ -bundle is the same as the data consisting of a line bundle η and a vector bundle V of rank 4 with $\eta^2 \approx \det V$. Also $Nm(\eta, V) = \eta$ so that (η, V) is of odd type if and only if V has degree $\equiv 2 \pmod{4}$. If α is any line bundle, $\alpha \circ (\eta, V) = (\alpha^2 \otimes \eta, \alpha \otimes V)$. Two $\Gamma^+(6)$ -bundles $(\eta, V), (\eta', V')$ are isomorphic as $\Gamma(6)$ -bundles if and only if $\eta = \eta'$ and $V = V'$ or $\eta' = \eta^{-1}$ and $V' = \eta^{-1} \otimes V^*$. Hence an isomorphism $(\eta, V) \approx i^*(\eta, V) \circ Nm(\eta, V) \circ \beta^{-1}$ for some β with $\beta^2 = h^{-\deg V}$ yields $V \approx i^*V \otimes \eta \otimes \beta^{-1}$, or $V \approx i^*V^* \otimes \beta^{-1}$. Thus we deduce: Consider stable vector bundles of rank 4 and fixed determinant of degree 2. Assume given an isomorphism $i^*V \approx V^* \otimes \beta^{-1}$ with $\beta^2 = h$. The moduli of such objects can thus be described as the variety of subspaces isotropic for a pencil of quadrics.

Remark 6.5. The constructions involved in the proof of the main theorems seem to be of a very general nature. For one thing, one can obtain results similar to [1, Theorem 3] for spin bundles with an i -action. Furthermore, it is also possible to get results for symplectic bundles, etc. However one ought to be able to understand these results in a group scheme theoretic set-up. In other words, given a bundle with an i -action, we modify the bundle at the Weierstrass points so that it goes down to $\mathbf{P}^1$. By keeping track of the additional structures on the quotient bundle, one would expect to classify bundles on X with i -action.

References

- [1] Usha V Desale and Ramanan S 1976 Classification of vector bundles of rank 2 on hyper-elliptic curves; *Inventiones Math.* **38** 161-185
- [2] Luc Gauthier 1954-55 Footnote to a footnote of André Weil. *Univ. de Politec. Torino Rend. Sem. Math.* **14** 325-328.
- [3] Mumford D 1965 *Geometric invariant theory*, (Springer-Verlag)
- [4] Narasimhan M S and Ramanan S 1975 Generalised Prym varieties as fixed points; *J. Indian Math. Soc.* p. 1-19
- [5] Ramanathan A, Stable principal bundles on a compact Riemann surface, Springer Lecture Notes (to appear)
- [6] Reid M *Intersection of two or more quadrics*, Thesis, Cambridge
- [7] Serre J P *Cohomologie Galoisienne*, (Paris : Hermann)
- [8] Seshadri C S 1967 Space of unitary vector bundles on a compact Riemann surface; *Ann. Math.* **85** 303-336.
- [9] A Weil 1954 Footnote to a recent paper; *Am. J. Math.* **76** 347-350

Institute for Advanced
Study,
Princeton, New Jersey, USA

and

Tata Institute of Fundamental
Research,
Bombay 400 005

EOL and ETOL array languages

NALINAKSHI NIRMAL and KAMALA KRITHIVASAN*

Department of Mathematics, Madras Christian College, Madras 600 059, India

* Computer Centre, Indian Institute of Technology, Madras 600 036, India

MS received 19 April 1980

Abstract. In this paper we study the families of ETOL and EOL array languages. Standard forms for ETOL and EOL array systems are defined and closure properties of the families are studied. Relations of these families with other developmental array languages and other array languages are studied.

Keywords. EOL array systems ; ETOL array systems ; normal form ; synchronized form ; closure properties ; hierarchy.

1. Introduction

In the theory of formal languages two generation mechanisms among others have been investigated in the past : Chomsky grammars and Lindenmayer-systems, particularly the latter field, has become a research area of enormous interest in the last few years [3, 5, 6].

Renewed interest has been evident in the last twelve years in extending the above two generation mechanisms to two dimensions [1, 2, 7 to 11]. However, all these constructs were such that the graphs or arrays were either allowed to grow only at the edges or surfaces or substitution for nodes were allowed only sequentially. But for biological reasons simultaneous rewriting is required and it should be possible to add new cells in the interior of the array.

In [1] cells divide simultaneously into at most two cells and new structures are defined by orienting the newly formed boundaries according to the configuration of all the neighbouring cells. In [2] there is an extension of parallel rewriting on string of symbols to graph generating systems.

In [4], we have defined a new array model in which parallel rewriting of every symbol in a rectangular array is considered where each symbol is replaced by an array of the same size and where we do not make any distinction between terminals and nonterminals. The language generated by it is the set of all arrays derived from the axiom, including the axiom, where the axiom is a rectangular array.

In this paper we generalize the notion of TOL array systems by adding an extra component called the *target* alphabet which is a subset of the total alphabet of a system and the language generated by it is defined as the set of all arrays over the

target alphabet which can be derived from the axiom. It is interesting to note that unlike the string languages [3], the ETOL array languages are not closed under row catenation, row +, column catenation, column +, like the other array languages [9, 10, 11]. ETOL array languages and EOL array languages are closed under pictorial transformations such as reflection about the right most vertical, half turn, reflection about the base, transpose and conjugation.

The paper is divided into five sections. In the next section we define all necessary concepts and investigate some basic properties of ETOLAS and EOLAS. In § 3, we search for sub-classes of the class of ETOLAS which are powerful enough to generate the whole class of ETOLAL. In § 4, we consider the closure and non closure properties of the ETOLAL and EOLAL. It is interesting to observe that ETOLAL and EOLAL are closed only under union and array homomorphism whereas a subfamily DEOLAL is not closed even under these two operations. In § 5, we compare different subfamilies of ETOLAL and compare them with the matrix models and Kolam arrays.

2. Definitions and basic properties

In this section, we review some basic definitions needed for this paper. For the definitions of AFM and array languages the reader is referred to [9, 10].

Notation : Let I be an alphabet, a finite nonempty set of symbols. A matrix (or array) over I is an $m \times n$ rectangular array of symbols from I ($m, n \geq 1$). The set of all matrices over I (including the empty matrix $\wedge$) is denoted by I^{**} and $I^{++} = I^{**} - \{\wedge\}$.

Definition 2.1

$$\text{If } X = \begin{matrix} a_{11} \dots a_{1n} \\ \dots\dots\dots \\ a_{m1} \dots a_{mn} \end{matrix}, \quad Y = \begin{matrix} b_{11} \dots b_{1n'} \\ \dots\dots\dots \\ b_{m'1} \dots b_{m'n'} \end{matrix}$$

The column catenation $X \oplus Y$ is defined only when $m = m'$ and is given by

$$\begin{matrix} a_{11} \dots a_{1n} b_{11} \dots b_{1n'} \\ \dots\dots\dots \\ a_{m1} \dots a_{mn} b_{m1} \dots b_{mn'} \end{matrix}$$

and the row catenation $X \ominus Y$ is defined only when $n = n'$ and is given by

$$\begin{matrix} a_{11} \dots a_{1n} \\ \dots\dots\dots \\ a_{m1} \dots a_{mn} \end{matrix}$$

$$b_{11} \dots b_{1n}$$

$$\dots\dots\dots$$

$$\dots\dots\dots$$

$$b_{m'1} \dots b_{m'n}$$

When there is no ambiguity and the meaning is clear $\oplus$ or $\ominus$ is left out.

Definition 2.2. If M and M' are two sets of matrices, the column product $M \oplus M' = \{X \oplus Y | X \text{ in } M, Y \text{ in } M'\}$ and the row product $M \ominus M' = \{X \ominus Y | X \text{ in } M, Y \text{ in } M'\}$.

Definition 2.3. An extended OL-array system (EOLAS) is a 4-tuple $G = (V, P, \omega, \Sigma)$ where :

1. V is a finite set (called the *alphabet* of G),
2. $\Sigma \subseteq V$ is a finite set called *terminal* or *target* alphabet.
3. $\omega \in V^{++}$ called *axiom* of G .
4. P is finite set of pairs (a, x) with a in V and x in V^{**} of dimensions $r \times s$ such that for each a in V at least one such pair is in P . The pairs (a, x) are called the rules or productions and are written as $a \rightarrow x$ or $a \xrightarrow{P} x$.

The elements of $V - \Sigma$ are called *auxiliary symbols* or *nonterminals*. EOLAS is called a *OLAS* if $V = \Sigma$.

The EOLAS is called *deterministic* if for every $a \in V$ there is exactly one a in V^{**} such that $(a, a) \in P$ and the array system is denoted by DEOLAS. EOLAS is called *propagating* if there is no rule of the type $a \rightarrow \lambda$ in P .

Definition 2.4. An extended tabled OL-array system (ETOLAS) is a 4-tuple $G = (V, \mathcal{P}, \omega, \Sigma)$ such that V, Σ, ω are same as that in EOLAS. $\mathcal{P}$ consists of a finite set $\{P_1, \dots, P_f\}$ for some $f \geq 1$ and each P_i is a finite subset of $V \times V^{**}$ with the following two conditions :

1. $(\forall_P)_{\mathcal{P}} (\forall_a)_V (\exists \alpha)_{V^{**}} ((a, \alpha) \in P)$,
2. $(\forall \langle a, \alpha \rangle)_P$, α 's are of the same size.

ETOLAS is called *deterministic* if for each P_i and each a in V there is exactly one α in V^{++} such that $(a, \alpha) \in P_i$ and it is denoted by DETOLAS. ETOLAS is called *propagating* if there is no rule of the type $a \rightarrow \lambda$ in any P in $\mathcal{P}$ and it is denoted by PETOLAS. ETOLAS is called a *TOLAS* if $V = \Sigma$.

Definition 2.5. Let $G = (V, \mathcal{P}, \omega, \Sigma)$ be an ETOLAS (EOLAS).

	$a_{11} \dots a_{1n}$	$M_{11} \dots M_{1n}$
	$\dots\dots\dots$	$\dots\dots\dots$
Let $u =$	$\dots\dots\dots$	$\dots\dots\dots$
	$a_{m1} \dots a_{mn}$	$M_{m1} \dots M_{mn}$

and $v =$

$a_{ij} \in V$, $M_{ij} \in V^{**}$, and M_{ij} , $1 \leq i \leq m$, $1 \leq j \leq n$ are arrays of the same dimension. Then we say $u \Rightarrow v$ iff there exists a table P_i such that $a_{ij} \rightarrow M_{ij}$ are in P_i . $\Rightarrow_G^*$ denotes the reflexive transitive closure of $\Rightarrow_G$.

Definition 2.6. Let $G = (V, \mathcal{P}, \omega, \Sigma)$ be an ETOLAS (EOLAS). The language generated by G is defined as

$$L(G) = \{x \in \Sigma^{++} / \omega \Rightarrow_G^* x\}.$$

A language $L \subseteq \Sigma^{**}$ is called a ETOLAL (EOLAL) iff there exists a ETOLAS (EOLAS) G such that $L = L(G)$. The family of ETOLAL (EOLAL) is denoted by $\mathcal{ETOLAL}$ ($\mathcal{EOLAL}$).

Now let us study a few basic properties of ETOLAS (EOLAS) and ETOL array languages (EOL array languages). We just state the first three lemmas as the proofs are simple.

Lemma 2.1. If $G = (V, \mathcal{P}, \omega, \Sigma)$ is an ETOLAS, then $L(G) = L(H) \cap \Sigma^{**}$, where $H = (V, \mathcal{P}, \omega, V)$ is a TOLAS.

Lemma 2.2. If G is an ETOLAS (EOLAS), we can construct an equivalent ETOLAS (EOLAS) $H = (V, \mathcal{P}, \omega, \Sigma)$ such that $\omega \in V - \Sigma$. From now onwards we can take the axiom to be a single letter $S \in V - \Sigma$.

Lemma 2.3. There is an algorithm which for every ETOLAS (EOLAS) G constructs an ETOLAS (EOLAS) H such that $L(H) = L(G) \cup \{\lambda\}$.

Lemma 2.4. There exists an algorithm which for every ETOLAS (EOLAS) G and every coding h produces an ETOLAS (EOLAS) H such that $L(H) = h(L(G))$.

Proof. We shall consider ETOLAS and EOLAS separately.

(i) Let $G = (V, \mathcal{P}, S, \Sigma)$ be an ETOLAS and let h be a coding from Σ into Σ_1 where we can take $V \cap \Sigma_1 = \emptyset$.

Let $H = (V \cup \Sigma_1 \cup \{F\}, \bar{\mathcal{P}} \cup P_0, S, \Sigma_1)$ be an ETOLAS such that $F \notin V \cup \Sigma_1$ and

$$P_0 = \{a \rightarrow h(a) : a \in \Sigma \text{ and } h(a) \text{ is defined}\} \\ \{a \rightarrow F : a \in \Sigma \text{ and } h(a) \text{ is not defined or } a \notin \Sigma\}.$$

$\bar{P}_i = P_i \cup \{a \rightarrow M'_{r_i s_i} : a \notin V, M'_{r_i s_i} \text{ is an array of } F\text{'s of dimension } r_i \times s_i\}$ and $\bar{\mathcal{P}} = \{\bar{P} / P \in \mathcal{P}\}$.

By the construction of H , it is clear that $L(H) = h(L(G))$.

(ii) Let $G = (V, P, S, \Sigma)$ be an EOLAS and let h be a coding from Σ into Σ_1 , where $V \cap \Sigma_1 = \emptyset$.

Let $H = (V \cup \Sigma_1 \cup \{F\}, \bar{P}, S, \Sigma_1)$ be an EOLAS such that $F \notin V \cup \Sigma_1$.

Let $\bar{P} = P \cup \{a \rightarrow h(a) : a \in \Sigma \text{ and } h(a) \text{ is defined}\} \cup$
 $\{a \rightarrow F : (a \in \Sigma \text{ and } h(a) \text{ is not defined}) \text{ or } a \notin \Sigma\}$
 $\cup \{a \rightarrow M'_{r_i s_i}, F \rightarrow M'_{r_i s_i}, a \in \Sigma_1\},$

where $M'_{r_i s_i}$ are arrays of F of dimension $r_i \times s_i$ and for every $a \in V$ we have a rule $\{a \rightarrow M_{r_i s_i} \text{ in } P\}$. Then clearly $L(H) = h(L(G))$.

Lemma 2.5. Every finite matrix language is an EOL array language.

Proof. Let $L = \{M_1, M_2, \dots, M_k\}$ be a finite matrix language over the alphabet Σ . Let S and F be not in Σ . Now consider a EOLAS $G = (\Sigma \cup \{S, F\}, P, S, \Sigma)$, where $P = \{S \rightarrow M_i : 1 \leq i \leq k\} \cup \{a \rightarrow M'_i : 1 \leq i \leq k, a \in \Sigma \cup \{F\}\}$, where M'_i is an array of F of dimension equal to that of M_i . Then clearly $L = L(G)$.

Let us prove the membership problem of OLAS, EOLAS, TOLAS and ETOLAS.

Definition 2.7. For a OLAS $G = (\Sigma, P, \omega)$ and for any non-negative integer n , we define the finite matrix language $L_n(G)$ by induction on n . $L_0(G) = \{\omega\}$, $L_{n+1}(G) = \{y / (\exists z) (z \in L_n(G) \text{ and } z \Rightarrow y, y \in \Sigma^{++})\}$.

Definition 2.8. For any TOLAS $G = (\Sigma, \mathcal{P}, \omega)$, for any non-negative integer n , we define the finite language $L_n(G)$ by induction on n . $L_0(G) = \{\omega\}$, $L_{n+1}(G) = \{y / \text{there exists a } z \text{ in } L_n(G) \text{ and a } P \text{ in } \mathcal{P} \text{ such that } z \Rightarrow_P y, y \in \Sigma^{++}\}$.

Lemma 2.6. The membership problem for OLAS (TOLAS) is solvable.

Let $G = (\Sigma, P, \omega)$ be the given OLAS and let M be the given array. Define for all n , $K_n(G) = \{a / a \in L_n(G) \text{ for some } m \in \{0, 1, \dots, n\} \text{ and } |a| \leq |M|\}$, where $|a|$ denotes the dimension of the array and it is denoted by (r, s) . Then $K_{n+1}(G) = K_n(G) \cup \{a / (\exists y)_{K_n(G)} (y \Rightarrow a) \text{ and } |a| \leq |M|\}$. Hence we get finite sets $K_0(G), K_1(G), \dots$. We have finitely many arrays a in Σ^{++} such that $|a| \leq |M|$. So we find a r such that $K_m(G) = K_r(G)$ for all $m \geq r$. Therefore $M \in L(G)$ iff $M \in \bigcup_{i=1}^r K_i(G)$. Hence by the above procedure we decide in a finite number of steps whether or not $M \in L(G)$.

Theorem 2.1. There is an algorithm which, for any EOL (ETOL) array system G and any array M , decides whether or not M is in $L(G)$.

Proof. We give the proof for EOLAS, proof for ETOLAS is analogous.

Let $G = (V, P, S, \Sigma)$ be an EOLAS. Let $\bar{G} = (V, P, S)$ be a OLAS. Then M is in $L(G)$ iff M is in $L(\bar{G})$ and M is in Σ^{++} . But it is trivial to decide whether or not M is in Σ^{++} . Hence the theorem follows from Lemma 2.6.

3. Normal forms of ETOLAS and EOLAS

In this section we search for subclasses of the class of ETOLAS (EOLAS) which are powerful enough to generate the entire class of ETOLAL (EOLAL),

Definition 3.1. An ETOLAS $G = (V, \mathcal{P}, S, \Sigma)$ is said to be in normal form if

1. $S \in V - \Sigma$ and S will not appear on the right side of any production in any table.
2. There exists a unique table P_I (called the initial table) such that $S \xrightarrow{P_I} a$ for some $a \in (V - \Sigma)^{++}$, $a \neq S$ and if $a \neq S$, the only production for a in P_I is a M'_{rs} where M'_{rs} is an array of F of dimension equal to that of a . (F is a distinguished symbol of $V - \Sigma$ called the rejection symbol. An array of F 's is called the rejection array).
3. There exists a unique table P_T (called the terminal table) different from P_I such that if $a \in V - \Sigma - \{S, F\}$ and $a \rightarrow a$ is in P_T then $a \in \Sigma \cup \{F\}$ and if $a \in \Sigma$ and $a \rightarrow a$ is in P_T , then $a = a$. If $a \in \{S, F\}$ then $a \rightarrow a$ implies $a = F$.
4. For all other tables $P_i \neq P_I, P_T$, $1 \leq i \leq n$, $a \rightarrow a$ is in P_i for all $a \in \Sigma$ then $a = M'_{rsi}$ where M'_{rsi} is the rejection array. If $a \rightarrow a$ is in P_i and $a \in V - \Sigma - \{S, F\}$ then $a \in (V - \Sigma)^{++}$. If $a \rightarrow a$ is in P_i and $a \in \{S, F\}$ then $a = M'_{rsi}$. The only production with F on the left side in every table is of the type $F \rightarrow M'_{rsi}$, where M'_{rsi} is the rejection array.

Theorem 3.1. There exists an algorithm which for every ETOLAS G produces an equivalent ETOLAS H which is in the normal form.

Proof. Let $G = (V, \mathcal{P}, \omega, \Sigma)$ be an ETOLAS. Let $G' = (V \cup \{S, F\}, \mathcal{P}', S, \Sigma)$ be an ETOLAS where $\{S, F\} \notin V$, and $\mathcal{P}' = P_0 \cup \{P'_i : P_i \in \mathcal{P}\}$, where $\# \mathcal{P} = n$. Let $P_0 = \{S \rightarrow \omega, a \rightarrow M'_{rs0} : a \in V \cup \{F\}\}$, $P'_i = P_i \cup \{S \rightarrow M'_{rs'i}, F \rightarrow M'_{rs'i}\}$, $1 \leq i \leq n$, where $M'_{rs'i}$ are rejection arrays. It is clear that $L(G) = L(G')$. Let $V_1 = \{\bar{a}/a \in V \cup \{S\}\} \cup \{F\}$, $P_1 = \{\bar{S} \rightarrow \bar{\omega} : S \rightarrow \omega \in P_0\} \cup \{a \rightarrow M'_{rs0} : a \in (V_1 - \{\bar{S}\}) \cup \Sigma\}$. $\bar{P}'_i = \{\bar{a} \rightarrow \bar{a} : a \rightarrow a \in P_i\} \cup \{\bar{S} \rightarrow M'_{rs'i}, F \rightarrow M'_{rs'i}\} \cup \{a \rightarrow M'_{rs'i} : a \in \Sigma\}$, $1 \leq i \leq n$, and $P_T = \{\bar{a} \rightarrow a : \bar{a} \in \Sigma\} \cup \{a \rightarrow a : a \in \Sigma\} \cup \{a \rightarrow F : a \notin \bar{\Sigma} \cup \Sigma\}$. Then $H = (V_1 \cup \Sigma, \bar{\mathcal{P}}, \bar{S}, \Sigma)$ is an ETOLAS which satisfies all the requirements for H to be normal where $\bar{\mathcal{P}} = \{P_1, P_T\} \cup \{\bar{P}'_i : P_i \in \mathcal{P}\}$. It follows from construction that $L(G) = L(H)$.

After constructing the normal form of an ETOLAS, we now construct a synchronized ETOLAS and synchronized EOLAS.

Definition 3.2. An ETOLAS (EOLAS) $G = (V, \mathcal{P}, S, \Sigma)$ ($G = (V, P, S, \Sigma)$) is said to be synchronized iff it has the following property : For any $M_1, M_2 \in V^{++}$ such that $M_1 \Rightarrow M_2$ and $M_1 \neq M_2$, if $M_1 \notin (V - \Sigma)^{++}$ then $M_2 \notin \Sigma^{++}$ that is, if an array contains an element of Σ , then all other arrays derivable from it contain some element not in Σ .

Theorem 3.2. For every ETOLAS there exists an equivalent synchronized ETOLAS.

Proof. Let $G = (V, \mathcal{P}, S, \Sigma)$ be any ETOLAS. By theorem 3.1 we may assume that G is in the normal form. In the terminal table P_T of G replace each produc-

tion of the type $a \rightarrow a$ where $a \in \Sigma$, by the production $a \rightarrow F$, where F is the rejection symbol. Then the system so obtained is the synchronized ETOL array system. Hence the theorem.

Theorem 3.3. There is an algorithm which for any EOLAS G produces a synchronized EOLAS $\bar{G}$ such that $L(G) = L(\bar{G})$.

Proof. Let $G = (V, P, \omega, \Sigma)$ be an EOLAS. Let $\bar{\Sigma} = \{\bar{a}/a \in \Sigma\}$, $(V - \Sigma) \cap \bar{\Sigma} = \emptyset$. If $a \neq b$ then $\bar{a} \neq \bar{b}$ in $\bar{\Sigma}$. Let $\bar{V} = V \cup \bar{\Sigma} \cup \{F, S\}$ where $S, F \notin V - \bar{\Sigma}$. If a is in V^{++} , we associate the array $\bar{a}$ in $(\bar{V})^{++}$ as follows :

$$\begin{array}{ccc} & \alpha_{11} \dots \alpha_{1n} & \beta_{11} \dots \beta_{1n} \\ \text{If } a = & \dots & \text{then } \bar{a} = \dots \text{ where} \\ & \alpha_{m1} \dots \alpha_{mn} & \beta_{m1} \dots \beta_{mn} \end{array}$$

$\beta_{ij} = \bar{a}_{ij}$ if a_{ij} is in Σ and $\beta_{ij} = a_{ij}$ if a_{ij} is not in Σ . Let $\bar{G} = (\bar{V}, \bar{P}, S, \Sigma)$ be an EOLAS where $\bar{P} = \{a \rightarrow \bar{a}/a \in V - \Sigma, a \rightarrow a \text{ in } P\} \cup \{a \rightarrow a/a \in V - \Sigma, a \rightarrow a \text{ in } P\} \cup \{\bar{a} \rightarrow \bar{a}/a \in \Sigma, a \rightarrow a \text{ in } P\} \cup \{a \rightarrow a'/a \in \Sigma \cup \{F\}, a \rightarrow a \text{ in } P \text{ where } a' \text{ is the rejection array where dimension is equal to that of } a\} \cup \{S \rightarrow \omega, S \rightarrow \bar{\omega}\} \cup \{a \rightarrow a'/a \in \bar{V} - \{S\}, a' \text{ is the rejection array where } |a| = |\omega|\}$. Then clearly $\bar{G}$ is synchronized EOLAS and $L(G) = L(\bar{G})$.

4. Closure properties of the class of ETOLAL and EOLAL

In this section we consider a number of operations like union, row catenation, column catenation, row +, column +, array homomorphism, complementation, on the class of ETOLAL and EOLAL. In string languages the family of EOL languages and the family of ETOL languages form an abstract family of languages [3]. But in array languages, the class of ETOLAL and the class of EOLAL are not closed under row catenation, row +, column catenation, column +.

Like the other known family of array languages [4, 9, 10], the family of ETOLAL and family of EOLAL are closed under quarter-turn, transpose, half turn, reflection about the rightmost vertical, reflection about the base and conjugation. For basic definitions refer [10].

Theorem 4.1. The family of ETOLAL (EOLAL) is closed under union and array homomorphism.

Proof. (i) *Union.* Let $G_1 = (V_1, \mathcal{P}_1, S_1, \Sigma_1)$ and $G_2 = (V_2, \mathcal{P}_2, S_2, \Sigma_2)$ be any two ETOL array systems. Without loss of generality let us assume G_1 and G_2 to be in normal form and that $(V_1 - (\Sigma_1 \cap \Sigma_2) - F) \cap (V_2 - (\Sigma_1 \cap \Sigma_2) - F) = \emptyset$. Let S be a new symbol such that $S \notin V_1 \cup V_2$. Let $P_0 = \{S \rightarrow S_1, S \rightarrow S_2\} \cup \{a \rightarrow F/a \in V_1 \cup V_2\}$. If $P_j \in \mathcal{P}_1$, then $\bar{P}_j = P_j \cup \{a \rightarrow M'/a \in V_2 \cup \{S\}\}$. If $P'_j \in \mathcal{P}_2$, then $\bar{P}'_j = \{a \rightarrow M''/a \in V_1 \cup \{S\}\} \cup P'_j$, where M' (M'') are rejection arrays whose dimensions equal the dimension of the right side of P_j (P'_j).

Let $G_3 = (V_1 \cup V_2 \cup \{S\}, \mathcal{P}_2, S, \Sigma_1 \cup \Sigma_2)$ be an ETOLAS where $\mathcal{P}_2 = P_0 \cup \{\bar{P}_j/P_j \in \mathcal{P}_1\} \cup \{\bar{P}'_j/P'_j \in \mathcal{P}'_1\}$. Then clearly $L(G_3) = L(G_1) \cup L(G_2)$;

(ii) *Array homomorphism.* Let $G = (V, \mathcal{P}, S, \Sigma)$ be an ETOLAS. Let $\Sigma = \{a_1, a_2, \dots, a_n\}$. Define an array homomorphism H from Σ into Δ as :

$$b_{11} \dots b_{1n}$$

$$H(a) = \dots \in \Delta^{++} \text{ and } a \in \Sigma. \text{ Let } P_0 = \{a \rightarrow H(a) \mid a \in \Sigma\},$$

$$b_{m1} \dots b_{mn}$$

$h(a) \in \Delta^{++} \cup \{b \rightarrow M'_{mn} \mid b \in \Delta \cup \{F\} \cup (V - \Sigma), |M'_{mn}| = |H(a)| \text{ and } M'_{mn} \text{ is the rejection array}\}$. Let $\bar{P}_{rj} = P_{rj} \cup \{b \rightarrow M' \mid b \in \Delta \cup \{F\}, M' \in \{F\}^{++}, |M'| \text{ is equal to the dimension of the right side of the rule of } P_{rj}\}$. Let $G_1 = (V \cup \Delta \cup \{F\}, \bar{\mathcal{P}}, S, \Delta)$ be an ETOLAS where $\bar{\mathcal{P}} = P_0 \cup \{\bar{P}_{rj} \mid P_{rj} \in \mathcal{P}_1\}$. Then clearly $L(G_1) = H(L(G))$.

Proof for showing that the $\mathcal{F}$ EOLAL is closed under union and homomorphism is analogous.

Theorem 4.2. The family of ETOLAL (EOLAL) is not closed under row catenation, column catenation, row +, column +, complementation.

Proof. (i) *Row Catenation* : Let $G_1 = (V_1, P_1, S_1, \Sigma_1)$ and $G_2 = (V_2, P_2, S_2, \Sigma_2)$ be two ETOLAS, where

$$P_1 = \{S_1 \rightarrow S_1 S_1, S_1 \rightarrow aa, a \rightarrow aa\} \text{ and}$$

$$P_2 = \left\{ S_2 \rightarrow \begin{matrix} S_2 S_2 \\ S_2 S_2 \end{matrix}, S_2 \rightarrow \begin{matrix} bb \\ bb \end{matrix}, b \rightarrow \begin{matrix} bb \\ bb \end{matrix} \right\}$$

Then it is clear that

$$\begin{matrix} L(G_1) \\ L(G_2) \end{matrix} = \left\{ \begin{matrix} aa & aaaa \\ bb & bbbb \\ bb & bbbb \end{matrix}, \dots \right\}$$

cannot be generated by any ETOLAS.

(ii) *Column catenations* : Let $G_3 = (V_3, P_3, S_3, \Sigma_3)$ be an ETOLAS, where

$$P_3 = \left\{ S_3 \rightarrow \begin{matrix} S_3 \\ S_3 \end{matrix}, S_3 \rightarrow \begin{matrix} a \\ a \end{matrix}, a \rightarrow \begin{matrix} a \\ a \end{matrix} \right\}.$$

Then it is clear that

$$L(G_3) \cdot L(G_2) = \left\{ \begin{matrix} abbbb \\ abb & abbbb \\ abb & abbbb \\ & abbbb \end{matrix}, \dots \right\}$$

cannot be generated by any ETOLAS.

(iii) *Row +* : We can show that $(L(G_3))_+$ cannot be generated by any ETOLAS.

(iv) *Column* : We can show that $(L(G_1))^+$ cannot be generated by any ETOLAS.

(v) *Complementation* : It is clear that the complement of $L(G_1)$ cannot be generated by any ETOLAS. Hence the theorem.

Theorem 4.3. The family of ETOLAL (EOLAL) is closed under quarter-turn, transpose, half-turn, reflection about the rightmost vertical, reflection about the base and conjugation.

Proof. Let $G = (V, \mathcal{P}, \omega, \Sigma)$ be a ETOLAS. Consider a ETOLAS $G_1 = (V, \mathcal{P}_1, \omega_1, \Sigma)$ where $\omega_1 = T(\omega)$, $T(A)$ denotes transpose of A . $\mathcal{P}_1 = \{P_1/P \in \mathcal{P}\}$ and $P_1 = \{a \rightarrow T(a)/a \rightarrow a \text{ in } P\}$. Then clearly $L(G_1) = T(L(G))$.

The proof for the other operations is analogous.

We observe that the family of EOLAL is closed under union and homomorphism, but it is of interest to note that a subfamily DEOLAL is not closed under these two operations. Before proving the nonclosure of $\mathcal{F}$ DEOLAL under union and array homomorphism, let us prove the following two lemmas.

Lemma 4.1. Let $a, b \in \Sigma$ and $a \neq b$. (i) If L is a DEOLAL then $\left\{ \begin{smallmatrix} aa & ab & ba \\ aa' & ab' & ba' \end{smallmatrix} \right\}$ is not a subset of L ; (ii) $\left\{ a, \begin{smallmatrix} aa \\ aa' \end{smallmatrix} \right\}$ is not a DEOLAL.

Proof. Let $G = (V, P, S, \Sigma)$ be a DEOLAS. Define a mapping μ from V^{++} into V^{++} as follows : $\mu^0(a) = a$, $\mu^1(a) = x$ if $a \rightarrow x$. For $n \geq 1$, $\mu^{n+1}(a) = \mu^1(\mu^n(a))$. We also define for $n \geq 0$, $m \geq 0$, $\mu^{n+m}(a) = \mu^n(a)$. $\mu^m(a)$, $\mu^n(ab) = \mu^n(a) \cdot \mu^n(b)$,

$$\mu^n \left(\begin{smallmatrix} a \\ \beta \end{smallmatrix} \right) = \begin{smallmatrix} \mu^n(a) \\ \mu^n(\beta) \end{smallmatrix} = \mu^n(a) \ominus \mu^n(\beta)$$

where $a, \beta \in V^{++}$. Let us assume that

$$\left\{ \begin{smallmatrix} aa & ab & ba \\ aa' & ab' & ba' \end{smallmatrix} \right\} \subseteq L(G),$$

i.e. we find i, j and k such that

$$S \xRightarrow{i} \begin{smallmatrix} aa \\ aa' \end{smallmatrix} \text{ i.e., } \mu^i(S) = \begin{smallmatrix} aa \\ aa' \end{smallmatrix}, \quad S \xRightarrow{j} \begin{smallmatrix} ab \\ ab' \end{smallmatrix} \text{ i.e., } \mu^j(S) = \begin{smallmatrix} ab \\ ab' \end{smallmatrix},$$

$$S \xRightarrow{k} \begin{smallmatrix} ba \\ ba' \end{smallmatrix} \text{ i.e., } \mu^k(S) = \begin{smallmatrix} ba \\ ba' \end{smallmatrix}.$$

Let us show that $i > j$. On the contrary let us take $i < j$ then

$$\begin{smallmatrix} ab \\ ab' \end{smallmatrix} = \mu^j(S) = \mu^{j-i}(\mu^i(S)) = \begin{smallmatrix} \mu^{j-i}(a) & \mu^{j-i}(a) \\ \mu^{j-i}(a) & \mu^{j-i}(a) \end{smallmatrix}.$$

Hence

$$a = \mu^{j-i}(a) \text{ and } b = \mu^{j-i}(a)$$

that is $a \xrightarrow{j-i} a$ and $a \xrightarrow{j-i} b$. This is not possible as the system is deterministic. Hence $i > j$. Similarly we show $i > k$. Next let us assume that $j > k$.

$$\frac{ab}{ab} = \mu^j(S) = \mu^{j-k}(\mu^k(S)) = \mu^{j-k} \begin{pmatrix} ba \\ ba \end{pmatrix} = \frac{\mu^{j-k}(b) \cdot \mu^{j-k}(a)}{\mu^{j-k}(b) \cdot \mu^{j-k}(a)}.$$

Hence $a = \mu^{j-k}(b)$ and $b = \mu^{j-k}(a)$ that is $b \xrightarrow{j-k} a$ and $a \xrightarrow{j-k} b$. Consider

$$\frac{aa}{aa} = \mu^i(S) = \mu^{i-k}(\mu^k(S)) = \mu^{i-k} \begin{pmatrix} ba \\ ba \end{pmatrix} = \frac{\mu^{i-k}(b) \cdot \mu^{i-k}(a)}{\mu^{i-k}(b) \cdot \mu^{i-k}(a)}.$$

Therefore $a = \mu^{i-k}(b)$ and $a = \mu^{i-k}(a)$. Hence $b \xrightarrow{i-k} a$ and $a \xrightarrow{i-k} a$.

Consider

$$\begin{aligned} \frac{aa}{aa} &= \frac{\mu^{i-k}(a) \cdot \mu^{i-k}(b)}{\mu^{i-k}(a) \cdot \mu^{i-k}(b)} = \frac{\mu^{i-k}(ab)}{\mu^{i-k}(ab)} = \mu^{i-k} \begin{pmatrix} ab \\ ab \end{pmatrix} = \mu^{i-k}(\mu^j(S)) \\ &= \mu^{i-k+j}(S) = \mu^{j-k}(\mu^i(S)) = \mu^{j-k} \begin{pmatrix} aa \\ aa \end{pmatrix} = \frac{\mu^{j-k}(a) \cdot \mu^{j-k}(a)}{\mu^{j-k}(a) \cdot \mu^{j-k}(a)}. \end{aligned}$$

Therefore $a = \mu^{j-k}(a)$ or $a \xrightarrow{j-k} a$. But we have already proved that $a \xrightarrow{j-k} b$. Hence a contradiction. Even if we take $j < k$, by a similar argument we arrive at a contradiction. Hence our assumption that $\left\{ \frac{aa}{aa}, \frac{ab}{ab}, \frac{ba}{ba} \right\} \subseteq L(G)$ is wrong.

So, $\left\{ \frac{aa}{aa}, \frac{ab}{ab}, \frac{ba}{ba} \right\}$ is not a subset of DEOLAL;

(ii) Let $G = (V, P, S, \Sigma)$ be a DEOLAS, so that $\left\{ a, \frac{aa}{aa} \right\} \subseteq L(G)$. Then

$$\begin{aligned} S \xRightarrow{i} a \text{ and } S \xRightarrow{j} \frac{aa}{aa}, \text{ i.e., } \mu^i(S) = a \text{ and } \mu^j(S) = \frac{aa}{aa}. \text{ If } j < i \text{ then } a = \mu^i(S) \\ = \mu^{i-j}(\mu^j(S)) = \mu^{i-j} \left(\frac{aa}{aa} \right) \text{ which is not possible. Hence } j > i. \text{ Now } \frac{aa}{aa} = \mu^j(S) \\ = \mu^{j-i}(\mu^i(S)) = \mu^{j-i}(a). \text{ Hence } a \xrightarrow{j-i} \frac{aa}{aa}. \end{aligned}$$

$$\begin{aligned} \text{Hence } \frac{aaaa}{aaaa} &= \frac{\mu^{j-i}(a) \cdot \mu^{j-i}(a)}{\mu^{j-i}(a) \cdot \mu^{j-i}(a)} = \frac{\mu^{j-i}(aa)}{\mu^{j-i}(aa)} = \mu^{j-i} \left(\frac{aa}{aa} \right) = \mu^{j-i}(\mu^j(S)) = \mu^{2j-i}(S). \end{aligned}$$

Therefore $S \xrightarrow{2j-i} (aaaa)_4$. Hence $(aaaa)_4 \notin L(G)$. So $\left\{ a, \frac{aa}{aa} \right\}$ is not a DEOLAL.

Lemma 4.2. Let $a, b \in \Sigma$ and $a \neq b$. If L is a DEOLAL then $\left\{ \frac{aa}{aa}, \frac{aa}{bb}, \frac{bb}{aa} \right\}$ is not a subset of L .

Proof. Proof is similar to lemma 4.1.

Theorem 4.4. DEOLAL is not closed under union and array homomorphism.

Proof. (i) *Union:* Let $G_1 = (V, P_1, \frac{ab}{ab}, V)$ and $G_2 = (\{a\}, P_2, \frac{aa}{aa}, \{a\})$ be two DEOLASs where $P_1 = \{a \rightarrow b, b \rightarrow a\}$ and $P_2 = \{a \rightarrow a\}$. Then $L(G_1) = \left\{ \frac{ab}{ab}, \frac{ba}{ba} \right\}$ and $L(G_2) = \left\{ \frac{aa}{aa} \right\}$. But $L(G_1) \cup L(G_2) = \left\{ \frac{ab}{ab}, \frac{ba}{ba}, \frac{aa}{aa} \right\}$ is not a DEOLAL follows from lemma 4.1.

(ii) *Array homomorphism:* Let $G_3 = (V, P_3, cd, V)$ be a DEOLAS where $P_3 = \{c \rightarrow a, d \rightarrow b, a \rightarrow b, b \rightarrow a\}$. Then $L(G_3) = \{cd, ab, ba\}$. Define a homomorphism h as $h(a) = h(c) = h(d) = \frac{a}{a}$, $h(b) = \frac{b}{b}$. Then

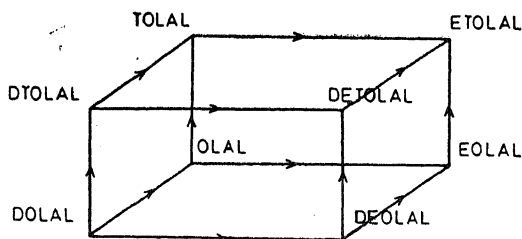
$$h(L(G_3)) = \left\{ \frac{aa}{aa}, \frac{ab}{ab}, \frac{ba}{ba} \right\}$$

which is not a DEOLAL follows from lemma 4.2.

5. Inter-relations among some subfamilies of ETOLAL

In this section, we investigate the hierarchy between the subfamilies of ETOLAL.

Theorem 5.1. The following diagram holds.



A solid line denotes strict inclusion in the direction indicated. If two families $\mathcal{F}_1$ and $\mathcal{F}_2$ are not connected by a path following the arrows in the diagram, then the families are incomparable but not disjoint.

Proof. Inclusions follow from definitions. For proper inclusions, refer to the examples in table 1. As the proof techniques are similar, we just prove that Example 3 in the table is an ETOLAL which cannot be generated by any DETOLAS.

Let $G = \left(\{a, b, c, d\}, \{P_1, P_2\}, \frac{aa}{bb}, \{a, b, c\} \right)$ be an ETOLAS, where $P_1 = \{a \rightarrow aa, b \rightarrow cc, c \rightarrow dd, b \rightarrow bb, d \rightarrow cc\}$, $P_2 = \{a \rightarrow ab, b \rightarrow ab, c \rightarrow dd, d \rightarrow cc\}$. Then

$$L(G) = \left\{ \frac{aa}{bb}, \frac{aaaa}{bbbb}, \frac{aaaa}{bbcc}, \frac{aaaa}{ccbb}, \frac{aaaa}{cccc}, \frac{aaaaaaaa}{ccbbccbb}, \dots \right\} = L.$$

Table 1.

Sl. No.	V	$\mathcal{P}$	Σ	ω	L in the family $\mathcal{F}_s$	L not in the family $\mathcal{F}_s$
1.	$\{a, b, c, d, e\}$	$\left\{ \begin{array}{l} a \rightarrow \frac{a}{a}, b \rightarrow \frac{b}{b}, c \rightarrow \frac{d}{d} \\ d \rightarrow \frac{e}{e}, e \rightarrow \frac{e}{e} \end{array} \right\}$	$\{a, b, c, d\}$	abc	DEOLAL	DOLAL, OLAL, DTOLAL, TOLAL
2.	$\{S, F, a\}$	$\left\{ \begin{array}{l} S \rightarrow a, a \rightarrow F, F \rightarrow F, \\ S \rightarrow \frac{aa}{aa}, a \rightarrow \frac{FF}{FF}, F \rightarrow \frac{FF}{FF} \end{array} \right\}$	$\{a\}$	S	EOLAL	DEOLAL, OLAL, TOLAL
3.	$\{a, b, c, d\}$	$\begin{array}{l} P_1 = (a \rightarrow aa, b \rightarrow cc, \\ c \rightarrow dd, b \rightarrow bb, d \rightarrow ee) \\ P_2 = (a \rightarrow ab, b \rightarrow ab, \\ c \rightarrow dd, d \rightarrow cc) \end{array}$	$\{a, b, c\}$	$\begin{array}{l} aa \\ bb \end{array}$	ETOLAL	EOLAL, DETOLAL, TOLAL
4.	$\{a, b, c, d\}$	$\begin{array}{l} P_1 = (a \rightarrow aa, b \rightarrow cc, c \rightarrow dd, d \rightarrow dd) \\ P_2 = (a \rightarrow ab, b \rightarrow bc, c \rightarrow dd, d \rightarrow dd) \end{array}$	$\{a, b, c\}$	abb	DETOLAL	EOLAL, TOLAL, DTOLAL, DEOLAL

If possible let there be a DETOLAS $G' = (V, \mathcal{P}, S, \{a, b, c\})$ so that $L(G') = L$. Without loss of generality, let us assume G' to be in the normal form. As $\begin{smallmatrix} aa \\ bb \end{smallmatrix}$

is the smallest array of L , we have $S \xRightarrow{*} \begin{smallmatrix} \bar{a} \bar{a} \\ \bar{b} \bar{b} \end{smallmatrix} \xRightarrow{P_T} \begin{smallmatrix} aa \\ bb \end{smallmatrix}$. Now $\begin{smallmatrix} \bar{a} \bar{a} \bar{a} \bar{a} \\ \bar{c} \bar{c} \bar{b} \bar{b} \end{smallmatrix}$ and $\begin{smallmatrix} \bar{a} \bar{a} \bar{a} \bar{a} \\ \bar{b} \bar{b} \bar{c} \bar{c} \end{smallmatrix}$

are not derivable from $\begin{smallmatrix} \bar{a} \bar{a} \\ \bar{b} \bar{b} \end{smallmatrix}$. Hence we have derivations

$$S \xRightarrow{*} \begin{smallmatrix} \bar{a} \bar{a} \bar{a} \bar{a} \\ \bar{c} \bar{c} \bar{b} \bar{b} \end{smallmatrix} \xRightarrow{P_T} \begin{smallmatrix} aaaa \\ cbbb \end{smallmatrix} \quad \text{and} \quad S \xRightarrow{*} \begin{smallmatrix} \bar{a} \bar{a} \bar{a} \bar{a} \\ \bar{b} \bar{b} \bar{c} \bar{c} \end{smallmatrix} \xRightarrow{P_T} \begin{smallmatrix} aaaa \\ bbcc \end{smallmatrix}.$$

where $V \supset \{\bar{a}, \bar{b}, \bar{c}\}$. If $S \xRightarrow{*}_{G'} M \xRightarrow{*}_{G'} \begin{smallmatrix} aaaaaaaaa \\ cbbccbb \end{smallmatrix}$, then

$$M \notin \left\{ \begin{smallmatrix} \bar{a} \bar{a} \bar{a} \bar{a} \\ \bar{b} \bar{b} \bar{b} \bar{b} \end{smallmatrix}, \begin{smallmatrix} \bar{a} \bar{a} \bar{a} \bar{a} \\ \bar{b} \bar{b} \bar{c} \bar{c} \end{smallmatrix}, \begin{smallmatrix} \bar{a} \bar{a} \bar{a} \bar{a} \\ \bar{c} \bar{c} \bar{b} \bar{b} \end{smallmatrix}, \begin{smallmatrix} \bar{a} \bar{a} \bar{a} \bar{a} \\ \bar{c} \bar{c} \bar{c} \bar{c} \end{smallmatrix} \right\}.$$

So we have to introduce one more table in $\mathcal{P}$. Proceeding in this way to generate the words of L , we require an infinite number of tables, a contradiction. Hence L is not a DETOLAL. L is also not an EOLAL or a TOLAL.

Theorem 5.2. $\mathcal{F}_{FML} \subseteq \mathcal{F}_{EOLAL}$.

Proof. Inclusion follows from lemma 2.5. For proper inclusion consider an EOL array system.

$$G = (\{S, A, a\}, \left\{ S \rightarrow \begin{smallmatrix} AA \\ AA \end{smallmatrix}, A \rightarrow \begin{smallmatrix} aa \\ aa \end{smallmatrix}, a \rightarrow \begin{smallmatrix} aa \\ aa \end{smallmatrix} \right\}, S, \{a\}).$$

Then $L(G)$ is not a FML.

In any infinite ETOLAL (EOLAL) the length and breadth of the array increase exponentially and not linearly. Whereas in the case of RML, $(R : X) \wedge L(X = R, CF, CS)$ of [9, 10] the length or the breadth of the array or both increase linearly. Hence we have the following theorem.

Theorem 5.3. ETOLAL (EOLAL) is incomparable but not disjoint with RML and $(R : X) \wedge L$.

References

- [1] Carlyle J W, Greibach S A and Paz A 1974 *Switching and automatic theory*, 15th Annual Symp. p. 1
- [2] Culik K II and Lindenmayer A 1976 *J. Gen. System. Theory* 3 53
- [3] Herman G T and Rozenberg G 1974 *Developmental systems and languages* (Amsterdam: North Holland)
- [4] Krithivasan K and Nirmal N 1980 *J. Indian Inst. Sci.* 62 (A) pp. 101-110

- [5] Lindenmayer A and Rozenberg G 1976 *Automata languages development* (Amsterdam : North Holland)
- [6] Rozenberg G and Salomaa A 1980 *Mathematical Theory of L systems* (New York : Academic Press)
- [7] Siromoney G and Siromoney R 1975 *Comput. Graph. Image Process* 4 361
- [8] Siromoney G and Siromoney R 1977 *Inf. Contr.* 35 119
- [9] Siromoney G, Siromoney R and Krithivasan K 1972 *Comput. Graph. Image Process* 12 284
- [10] Siromoney G, Siromoney R and Krithivasan K 1973 *Inf. Contr.* 22 447
- [11] Siromoney G, Siromoney R and Krithivasan K 1974 *Comput. Graph. Image Process* 3 63

On the relation of generalized Valiron summability to Cesàro summability

V K KRISHNAN

St. Thomas College, Trichur 680 001, India

MS received 3 May 1980

Abstract. A family (V_a^k) of summability methods, called generalized Valiron summability, is defined. The well-known summability methods (B_a, γ) , (E_p) , (T_a) , (S_β) and (V_a) are members of this family. In §3 some properties of the (B_a, γ) and (V_a^k) transforms are established. Following Satz II of Faulhaber (1956) it is proved that the members of the (V_a^k) family are all equivalent for sequences of finite order. This paper is a good illustration of the use of generalized Boral summability. The following theorem is established :

Theorem. If s_n ($n \geq 0$) is a real sequence satisfying

$$\lim_{\epsilon \rightarrow 0+} \liminf_{m \rightarrow \infty} \min_{m \leq n \leq m + \epsilon \sqrt{m}} \left(\frac{s_n - s_m}{m^\rho} \right) \geq 0 \quad (\rho \geq 0),$$

and if $s_n \rightarrow s(V_a^k)$ then $s_n \rightarrow s(C, 2\rho)$.

Keywords. Generalized Valiron summability ; Boral summability ; Rajagopal's theorem.

1. Introduction

Rajagopal ([4], Theorem 2) proved the following theorem connecting Borel and Cesàro summabilities ; and, after him, Sitaraman ([5], Theorem II) proved the theorem with Borel summability replaced by summability (S_β) defined as usual in § 5 :

Theorem A. If s_n ($n \geq 0$) is a real sequence satisfying

$$\lim_{\epsilon \rightarrow 0+} \liminf_{m \rightarrow \infty} \min_{m \leq n \leq m + \epsilon \sqrt{m}} \left(\frac{s_n - s_m}{m^\rho} \right) \geq 0 \quad (\rho \geq 0), \quad (1)$$

and if $s_n \rightarrow s(B)$, then $s_n \rightarrow s(C, 2\rho)$.

In this paper we prove (Theorem 4) that Theorem A is extensible to a family (V_a^k) of summability methods which include as special cases generalized Borel summability (B_a, γ) defined in § 2 and the well-known summabilities (E_p) , (T_a) , (S_β) defined in the usual notation in § 5. Of course Theorem A itself obviously

includes the similar theorem for summability (E_p) instead of summability (B) , since $(E_p) \subset (B)$. Valiron summability (V_a) is also a special case of summability (V_a^k) , as pointed out in § 5, and the latter is the generalized Valiron summability of the title.

The Tauberian condition (1) reduces to a classic special case when $\rho = 0$. A different special case of (1) and its further specialization are respectively

$$s_n - s_{n-1} = O_L(n^{\rho-1/2}), \quad (1a)$$

$$s_n - s_{n-1} = o(n^{\rho-1/2}). \quad (1b)$$

Hardy and Littlewood originally proved the special case of Theorem A with (1b) instead of (1), as stated by Hardy ([3], note on §§ 9.6-7). Their result was extended by Borwein [1] to generalized Borel summability, and an idea of his (Lemma 7) is used in the sequel.

2. Definitions

The V_a^k transform of a (generally complex) sequence s_n ($n \geq 0$) is the function defined by

$$V_a^k(x) = \sum_{n=0}^{\infty} c_n(x) s_n, \quad x > 0,$$

where $c_n(x)$ satisfies the following three conditions :

- (i) $c_n(x) \geq 0$ for $n = 0, 1, 2, \dots, x > 0$;
- (ii) there exist $a > 0$ and δ with $\frac{1}{2} < \delta < \frac{2}{3}$ such that, for every positive integer k , $c_n(x)$ can be expressed as

$$c_n(x) = \left(\frac{a}{\pi x}\right)^{1/2} \exp\{-ax^{-1}(n-x)^2 + g_k + R_k\}$$

whenever x is sufficiently large and $|n-x| \leq x^\delta$, and where

$$g_k = \sum_{i=0}^{2k-1} \sum_{j=0}^{i+1} l_{ij} \frac{(n-x)^j}{x^i}, \quad l_{12} = 0,$$

l_{ij} being independent of n and bounded as $x \rightarrow \infty$,

$$R_k = O\left(\frac{|n-x|^{2k+1} + 1}{x^{2k}}\right) \text{ as } x \rightarrow \infty$$

uniformly in n for $|n-x| \leq x^\delta$;

- (iii) for every $\sigma \geq 0$

$$\sum_{|n-x| > x^\delta} (n+1)^\sigma c_n(x) = o(1) \text{ as } x \rightarrow \infty.$$

We say that s_n is summable (V_a^k) to s (finite), and write $s_n \rightarrow s (V_a^k)$ if $V_a^k(x) \rightarrow s$ as $x \rightarrow \infty$.

The $(B_{a,\gamma})$ transform ($a > 0$, γ real) of s_n is the function defined by

$$B(x) = a \exp(-ax) \sum_{n=N}^{\infty} \frac{(ax)^{na+\gamma-1}}{\Gamma(na+\gamma)} s_n, \quad x > 0,$$

N being the least positive integer such that $Na + \gamma \geq 1$. We say that s_n is summable $(B_{a,\gamma})$ to s (finite), and write $s_n \rightarrow s(B_{a,\gamma})$ if $B(x) \rightarrow s$ as $x \rightarrow \infty$.

The n th Cesàro sum and the n th Cesàro mean of s_n , each of order $r > -1$, are denoted by S_n^r and s_n^r respectively. Thus

$$S_n^0 = s_n^0 = s_n; \quad S_n^r = \sum_{\nu=0}^n \binom{n-\nu+r-1}{n-\nu} s_\nu = s_n^r \binom{n+r}{n}.$$

We say that s_n is summable (C, r) to s (finite), and write $s_n \rightarrow s(C, r)$ if $s_n^r \rightarrow s$ as $n \rightarrow \infty$.

3. Preliminary results

In this section we study some properties of the $(B_{a,\gamma})$ and V_a^k transforms.

Lemma 1. The $(B_{a,\gamma})$ transform is a V_a^k transform with $a = a/2$.

Proof. Borwein ([1], Lemma 2 (d)) has proved that $c_n(x)$ defined as below satisfies condition (iii):

$$c_n(x) = a \exp(-ax) \frac{(ax)^{na+\gamma-1}}{\Gamma(na+\gamma)} \text{ for } n \geq N \text{ and } c_n(x) = 0 \text{ for } n < N.$$

To verify condition (ii), let $\frac{1}{2} < \delta < \frac{2}{3}$, x be large, $|n-x| \leq x^\delta$, and k be any positive integer. Writing $h = n - x + (\gamma-1)/a$ and using the formula

$$\begin{aligned} \log \Gamma(y+1) &= \frac{1}{2} \log(2\pi) + \left(y + \frac{1}{2}\right) \log y - y + \sum_{r=1}^k \frac{(-1)^{r-1} B_r}{(2r-1) 2r} y^{-2r+1} \\ &\quad + O(y^{-2k-1}) \text{ as } y \rightarrow \infty, \end{aligned}$$

with $y = ax + ah$ we see that

$$\begin{aligned} \log \left\{ c(x) \left(\frac{2\pi x}{a} \right)^{1/2} \right\} &= \frac{1}{2} \log(2\pi) - ax + \left(ax + ah + \frac{1}{2}\right) \log(ax) \\ &\quad - \log \Gamma(ax + ah + 1) \\ &= A_1 + A_2 + A_3 \end{aligned}$$

where

$$A_1 = ah + (ax + ah + \frac{1}{2}) \log(ax) - (ax + ah + \frac{1}{2}) \log(ax + ah), \quad (2)$$

$$A_2 = - \sum_{r=1}^k \frac{(-1)^{r-1} B_r}{(2r-1) 2r} (ax + ah)^{-2r+1}, \quad (3)$$

$$|A_3| \leq M(ax + ah)^{-2k-1} \quad (4)$$

for some constant M . By Taylor's theorem,

$$\log(1+y) = \sum_{r=1}^{2k} \frac{(-1)^{r-1}}{r} y^r + \frac{y^{2k+1}}{2k+1} (1+\theta y)^{-2k-1},$$

where $0 \leq \theta = \theta(k, y) \leq 1$. Therefore, from (2),

$$\begin{aligned} A_1 &= ah - \left(ax + ah + \frac{1}{2}\right) \log\left(1 + \frac{h}{x}\right) \\ &= ah - ax \left\{ \frac{h}{x} - \frac{1}{2} \left(\frac{h}{x}\right)^2 + \sum_{r=3}^{2k} \frac{(-1)^{r-1}}{r} \left(\frac{h}{x}\right)^r \right\} \\ &\quad - ah \left\{ \frac{h}{x} + \sum_{r=2}^{2k-1} \frac{(-1)^{r-1}}{r} \left(\frac{h}{x}\right)^r \right\} - \frac{1}{2} \sum_{r=1}^{2k-1} \frac{(-1)^{r-1}}{r} \left(\frac{h}{x}\right)^r + A_4 \\ &= -\frac{a}{2} \frac{h^2}{x} + \sum_{r=2}^{2k-1} u_r \frac{h^{r+1}}{x^r} + \sum_{r=1}^{2k-1} v_r \frac{h^r}{x^r} + A_4 \end{aligned}$$

where u_r, v_r are independent of h, x and

$$A_4 = \frac{a}{2k} \frac{h^{2k+1}}{x^{2k}} + \frac{1}{2} \frac{1}{2k} \frac{h^{2k}}{x^{2k}} - \frac{(ax + ah + \frac{1}{2})}{2k+1} \left(\frac{h}{x}\right)^{2k+1} \left(1 + \theta \frac{h}{x}\right)^{-2k-1}. \quad (5)$$

Again, Taylor's theorem gives

$$\begin{aligned} (1+y)^{-\mu} &= \sum_{v=0}^m (-1)^v \binom{\mu+v-1}{v} y^v + (-1)^{m+1} y^{m+1} \binom{\mu+m}{m+1} \\ &\quad \times (1+\theta y)^{-\mu-m-1}, \end{aligned}$$

where $0 \leq \theta = \theta(\mu, m, y) \leq 1$. Using this with $y = h/x$, $\mu = 2r-1$, $m = 2k-2r$, $r = 1, \dots, k$, we get from (3),

$$\begin{aligned} A_2 &= \sum_{r=1}^k \frac{(-1)^r B_r}{(2r-1)2r} \alpha^{-2r+1} x^{-2r+1} \left\{ \sum_{v=0}^{2k-2r} (-1)^v \binom{2r-1+v-1}{v} \left(\frac{h}{x}\right)^v \right. \\ &\quad \left. + (-1)^{2k-2r+1} \left(\frac{h}{x}\right)^{2k-2r+1} \binom{2k-1}{2k-2r+1} \left(1 + \theta_r \frac{h}{x}\right)^{-2k} \right\} \\ &= \sum_{r=1}^k \sum_{v=0}^{2k-2r} w_{r,v} \frac{h^v}{x^{2r+v-1}} + A_5, \end{aligned}$$

where $w_{r,v}$ are independent of h, x and

$$A_5 = \sum_{r=1}^k \frac{(-1)^{r+1} B_r}{(2r-1)2r} \alpha^{-2r+1} \frac{h^{2k-2r+1}}{x^{2k}} \binom{2k-1}{2k-2r+1} \left(1 + \theta_r \frac{h}{x}\right)^{-2k}. \quad (6)$$

Thus we have proved that

$$\log \left\{ c_n(x) \left(\frac{2\pi x}{a} \right)^{1/2} \right\} = -\frac{a}{2} \frac{h^2}{x} + \sum_{i=1}^{2k-1} \sum_{j=0}^{i+1} \bar{l}_{ij} \frac{h^j}{x^i} + A_3 + A_4 + A_5,$$

where $\bar{l}_{ij}$ are independent of h, x and $\bar{l}_{12} = 0$. Noting that $h = n - x + (\gamma - 1)/a$ and writing $a = 2a$ we get

$$\begin{aligned} \log c_n(x) &= \log \left(\frac{a}{\pi x} \right)^{1/2} - \frac{a}{x} \left\{ (n-x)^2 + 2(n-x) \left(\frac{\gamma-1}{a} \right) + \left(\frac{\gamma-1}{a} \right)^2 \right\} \\ &+ \sum_{i=1}^{2k-1} \sum_{j=0}^{i+1} \bar{l}_{ij} \frac{1}{x^i} \sum_{v=0}^j \binom{j}{v} (n-x)^v \left(\frac{\gamma-1}{a} \right)^{j-v} + A_3 + A_4 + A_5 \\ &= \log \left(\frac{a}{\pi x} \right)^{1/2} - ax^{-1}(n-x)^2 + \sum_{i=1}^{2k-1} \sum_{j=0}^{i+1} l_{ij} \frac{(n-x)^j}{x^i} + A_3 + A_4 + A_5, \end{aligned} \quad (7)$$

where l_{ij} are independent of n, x and $l_{12} = \bar{l}_{12} = 0$.

Since $h = n - x + (\gamma - 1)/a$, we have $|h|/x < \frac{1}{2}$ and $1 + \theta h/x > \frac{1}{2}$ whenever $0 \leq \theta \leq 1$, x is large and $|n - x| \leq x^\delta$. Moreover,

$$|h|^v \leq \{ |n - x|^{2k+1} + 1 \} \left\{ 1 + \frac{|\gamma - 1|}{a} \right\}^{2k+1}, \quad v = 0, 1, \dots, 2k+1.$$

Supplying these estimates in (4), (5) and (6), we find that

$$x^{2k} [|A_3| + |A_4| + |A_5|] \leq M' [|n - x|^{2k+1} + 1]$$

if $|n - x| \leq x^\delta$ and x is large, M' being a constant. This, in view of (7), completes the proof of the lemma.

Lemma 2. If $c_n(x)$ satisfies the conditions of a V_a^k transform, then for $\sigma \geq 0$,

$$(a) \quad c_n(x) = \{1 + \epsilon_1(n, x)\} \left(\frac{a}{\pi x} \right)^{1/2} \exp[-ax^{-1}(n-x)^2],$$

$$(b) \quad \left(\frac{n}{x} \right)^\sigma c_n(x) = \{1 + \epsilon_2(n, x)\} \int_n^{n+1} \left(\frac{a}{\pi x} \right)^{1/2} \exp[-ax^{-1}(t-x)^2] dt,$$

where $\epsilon_1(n, x), \epsilon_2(n, x) \rightarrow 0$ as $x \rightarrow \infty$ uniformly in n for $|n - x| \leq x^\delta$;

$$(c) \quad \sum_{n=1}^{\infty} \left(\frac{n}{x} \right)^\sigma c_n(x) \rightarrow 1 \text{ as } x \rightarrow \infty;$$

$$(d) \quad \theta_1(N, x) \equiv \sum_{n=N}^{\infty} \left(\frac{n}{x} \right)^\sigma c_n(x) (\sqrt{n} - \sqrt{N}) \rightarrow 0,$$

$$(e) \quad \theta_2(N, x) \equiv \sum_{n=N}^{\infty} \left(\frac{n}{x} \right)^\sigma c_n(x) \rightarrow 0,$$

$$(f) \theta_3(M, x) \equiv \sum_{n=1}^M \left(\frac{n}{x}\right)^\sigma c_n(x) \rightarrow 0,$$

as $x, N, \sqrt{N} - \sqrt{x}, M, \sqrt{x} - \sqrt{M} \rightarrow \infty$. (c) for $\sigma = 0$ shows that V_a^k is a positive regular transform of s_n .

Proof. (a) Taking $k = 1$ in condition (ii) on $c_n(x)$, we see that

$$\begin{aligned} \exp(g_1 + R_1) &= \exp\left\{\frac{l_{10}}{x} + l_{11} \frac{(n-x)}{x} + O\left(\frac{|n-x|^3 + 1}{x^2}\right)\right\} \\ &= 1 + O\left(\frac{1}{x} + \frac{|n-x|}{x} + \frac{|n-x|^3}{x^2}\right) \end{aligned}$$

which proves (a), since $\delta < 2/3$.

(b) Noting that $(n/x)^\sigma = 1 + \epsilon'(n, x)$ and that

$$\int_n^{n+1} \exp[-ax^{-1}(t-x)^2] dt = \exp[-ax^{-1}(n-x)^2] \{1 + \epsilon''(n, x)\}$$

we deduce (b) from (a).

(c) Write the sum in (c) as

$$\left(\sum_{|n-x| \leq x^\delta} + \sum_{|n-x| > x^\delta}\right) \left(\frac{n}{x}\right)^\sigma c_n(x) = S_1 + S_2.$$

From (b) it follows that

$$\begin{aligned} S_1 &= \int_{x-x^\delta}^{x+x^\delta} \left(\frac{a}{\pi x}\right)^{1/2} \exp\{-ax^{-1}(t-x)^2\} dt + o(1) \\ &= \int_{-x^{\delta-1/2}}^{x^{\delta-1/2}} \left(\frac{a}{\pi}\right)^{1/2} \exp(-au^2) du + o(1), \end{aligned}$$

which tends to 1 as $x \rightarrow \infty$ since $\delta > \frac{1}{2}$. On the other hand, $S_2 \rightarrow 0$ as $x \rightarrow \infty$ by our condition (iii) on $c_n(x)$.

(d) Write $\sqrt{N} - \sqrt{x} = u$ so that $N - x = u(\sqrt{N} + \sqrt{x}) > u\sqrt{x}$.

In view of condition (iii) on $c_n(x)$ it suffices to prove that

$$S = \sum_{N \leq n \leq x+x^\delta} \left(\frac{n}{x}\right)^\sigma c_n(x) (\sqrt{n} - \sqrt{N}) \rightarrow 0,$$

as $x, u \rightarrow \infty$. Since $\sqrt{n} - \sqrt{N} < (n-x)/\sqrt{x}$ for $n \geq N > x$, it follows from (b) that, for all large x ,

$$S \leq \sum_{N \leq n \leq x+x^\delta} \frac{n-x}{\sqrt{x}} 2 \int_n^{n+1} \left(\frac{a}{\pi x}\right)^{1/2} \exp[-ax^{-1}(t-x)^2] dt$$

$$\begin{aligned}
&\leq 2 \left(\frac{a}{\pi x} \right)^{1/2} \sum_{N \leq n \leq x+\delta} \int_n^{n+1} \frac{t-x}{\sqrt{x}} \exp[-ax^{-1}(t-x)^2] dt \\
&\leq 2 \left(\frac{a}{\pi x} \right)^{1/2} \int_N^\infty \frac{t-x}{\sqrt{x}} \exp[-ax^{-1}(t-x)^2] dt \\
&= 2 \left(\frac{a}{\pi} \right)^{1/2} \int_{\frac{N-x}{\sqrt{x}}}^\infty v \exp(-av^2) dv \\
&\leq 2 \left(\frac{a}{\pi} \right)^{1/2} \int_u^\infty v \exp(-av^2) dv
\end{aligned}$$

which tends to 0 as $u \rightarrow \infty$.

Proofs of (e) and (f) are similar to that of (d).

Lemma 3. If s_n is a real sequence satisfying (1) then there exist positive constants K_1, K_2 such that, for $n \geq m \geq 1$,

$$s_n - s_m > -K_1 n^\rho (\sqrt{n} - \sqrt{m}) - K_2 m^\rho.$$

This is proved exactly like Theorem 239 of Hardy [3] (see Rajagopal [4], Lemma 1).

Lemma 4. If s_n is a real sequence satisfying (1), and if

$$V_a^k(x) = O(x^\rho) \quad (x \rightarrow \infty), \quad (8)$$

then $s_n = O(n^\rho) \quad (n \rightarrow \infty)$.

Proof. It may be remarked that the proof is applicable to any positive regular transform, in place of $V_a^k(x)$, for which the $c_n(x)$ satisfy (c)–(f) of Lemma 2. We proceed on the lines of the proofs of Theorem 1 of Rajagopal [4] and Theorem 238 of Hardy [3]. Write, for $n \geq 1$

$$\sigma_n = \frac{s_n}{n^\rho}, \quad \sigma_1(n) = \max_{1 \leq r \leq n} \sigma_r, \quad \text{and} \quad \sigma_2(n) = \max_{1 \leq r \leq n} (-\sigma_r),$$

and assume, for convenience, that $s_0 = 0$. Since

$$x^{-\rho} V_a^k(x) = \sum_{n=1}^{\infty} (n/x)^\rho c_n(x) \sigma_n,$$

it follows from (8) and Lemma 2 (c) that neither $\sigma_n \rightarrow \infty$ nor $\sigma_n \rightarrow -\infty$ is possible. The lemma is proved by showing that each of the following two cases contradicts (8):

(I) $\sigma_1(n) \geq \sigma_2(n)$ for infinitely many n and $\sigma_1(n) \rightarrow \infty$;

(II) $\sigma_1(n) < \sigma_2(n)$ for all but a finite number of values of n and $\sigma_2(n) \rightarrow \infty$.

Case I. Corresponding to a large positive number H , choose the least $M = M(H)$ for which $\sigma_M = \sigma_1(M) > 2H$ and $\sigma_1(M) \geq \sigma_2(M)$, and then the

least $N = N(H) > M$ for which $\sigma_N \leq \frac{1}{2} \sigma_M$. Define $x = x(H)$ by $2\sqrt{x} = \sqrt{M} + \sqrt{N}$ and write

$$x^{-\rho} V_a^k(x) = \left(\sum_{n=1}^{M-1} + \sum_{n=M}^{N-1} + \sum_{n=N}^{\infty} \right) (n/x)^{\rho} c_n(x) \sigma_n = S_1 + S_2 + S_3. \quad (9)$$

If $(M/N)^{\rho} \leq 2/3$ then

$$\sqrt{N} - \sqrt{M} \geq \{1 - (2/3)^{1/2\rho}\} \sqrt{N}. \quad (10)$$

If $(M/N)^{\rho} > 2/3$ it follows from Lemma 3 that

$$\begin{aligned} \sigma_N - \sigma_M (M/N)^{\rho} &> -K_1(\sqrt{N} - \sqrt{M}) - K_2(M/N)^{\rho}, \\ K_1(\sqrt{N} - \sqrt{M}) &> -\sigma_N + \sigma_M (M/N)^{\rho} - K_2 \\ &> -\frac{1}{2}\sigma_M + \frac{2}{3}\sigma_M - K_2 \\ &> \frac{1}{3}H - K_2, \end{aligned}$$

by the choice of M, N . This, together with (10), shows that $\sqrt{N} - \sqrt{M} \rightarrow \infty$ as $H \rightarrow \infty$. Hence $\sqrt{N} - \sqrt{x}, \sqrt{x} - \sqrt{M} \rightarrow \infty$ as $H \rightarrow \infty$.

The estimates which follow are when $H \rightarrow \infty$ and so $x, M, N, \sqrt{x} - \sqrt{M}, \sqrt{N} - \sqrt{x} \rightarrow \infty$ as in Lemma 2. By the choice of N, M ,

$$s_{N-1} = (N-1)^{\rho} \sigma_{N-1} > (N-1)^{\rho} \frac{1}{2} \sigma_M > (N-1)^{\rho} H.$$

Hence, by Lemma 3, we have for $n \geq N$,

$$\begin{aligned} s_n &> -K_1 n^{\rho} \{\sqrt{n} - \sqrt{(N-1)}\} - K_2 (N-1)^{\rho} + s_{N-1} \\ &> -K_1 n^{\rho} \{\sqrt{n} - \sqrt{(N-1)}\}, \end{aligned}$$

and, therefore, by Lemma 2 (d), (e)

$$\begin{aligned} S_3 &> -K_1 \sum_{n=N}^{\infty} (n/x)^{\rho} c_n(x) \{\sqrt{n} - \sqrt{(N-1)}\} \\ &= -K_1 \theta_1(N-1, x). \end{aligned}$$

On the other hand, by Lemma 2 (c), (e), (f) and the choice of M, N ,

$$\begin{aligned} S_2 &\geq \frac{1}{2} \sigma_M \sum_{n=M}^{N-1} (n/x)^{\rho} c_n(x) = \frac{1}{2} \sigma_M \{1 + o(1) - \theta_2(N, x) - \theta_3(M-1, x)\}, \\ S_1 &\geq -\sigma_2(M) \sum_{n=1}^{M-1} (n/x)^{\rho} c_n(x) \geq -\sigma_1(M) \theta_3(M-1, x). \end{aligned}$$

Combining these estimates for S_1, S_2, S_3 we find from (9) that $x^{-\rho} V_a^k(x) \rightarrow \infty$ as $H \rightarrow \infty$ contradicting (8).

Case II. Corresponding to a large positive number H choose the least $N = N(H)$ such that $\sigma_2(n) > \sigma_1(n)$ for $n \geq N$ and $\sigma_N = -\sigma_2(N) < -2H$; and then the last $M = M(H) < N$ for which $\sigma_M \geq \frac{1}{2} \sigma_N = -\frac{1}{2} \sigma_2(N)$. Define x as in case I, and write

$$x^{-\rho} V_a^k(x) = \left(\sum_{n=1}^M + \sum_{n=M+1}^N + \sum_{n=N+1}^{\infty} \right) (n/x)^{\rho} c_n(x) \sigma_n = S_1 + S_2 + S_3. \quad (11)$$

Using Lemma 3 we find, as in case I, that

$$\begin{aligned} K_1(\sqrt{N} - \sqrt{M}) &> -\sigma_N + \sigma_M(M/N)^\rho - K_2(M/N)^\rho \\ &\geq -\sigma_N\{1 - \tfrac{1}{2}(M/N)^\rho\} - K_2 \\ &> H - K_2, \end{aligned}$$

and that $\sqrt{N} - \sqrt{x}$, $\sqrt{x} - \sqrt{M} \rightarrow \infty$ as $H \rightarrow \infty$.

From Lemma 3, it follows that, for $n \geq N$,

$$\begin{aligned} \sigma_n &> -\{-\sigma_N + K_2\}(N/n)^\rho - K_1(\sqrt{n} - \sqrt{N}) \\ &\geq -\{\sigma_2(N) + K_2\} - K_1(\sqrt{n} - \sqrt{N}) \\ &= -t_n, \end{aligned}$$

say. Thus $-\sigma_n < t_n$ for $n \geq N$; while $-\sigma_n \leq -\sigma_N < t_N$ for $n < N$ by our choice of N and definition of $\sigma_2(n)$. Since t_n is an increasing function of n , we thus have $-\sigma_m \leq t_m \leq t_n$ for $n \geq m \geq N$, and $-\sigma_m < t_N \leq t_n$ for $n \geq N > m$. This implies that $t_n \geq \sigma_2(n)$ for $n \geq N$ by the definition of $\sigma_2(n)$. Hence, by the choice of N and Lemma 2 (d), (e)

$$\sigma_n \leq \sigma_1(n) < \sigma_2(n) \leq t_n = \sigma_2(N) + K_2 + K_1(\sqrt{n} - \sqrt{N}) \quad (n \geq N),$$

$$\begin{aligned} S_3 &\leq \sum_{n=N+1}^{\infty} (n/x)^\rho c_n(x) \{\sigma_2(N) + K_2 + K_1(\sqrt{n} - \sqrt{N})\} \\ &\leq \{\sigma_2(N) + K_2\} \theta_2(N, x) + K_1 \theta_1(N, x). \end{aligned}$$

On the other hand, by Lemma 2 (c), (e), (f) and by the choice of M , N ,

$$\begin{aligned} S_2 &\leq -\tfrac{1}{2}\sigma_2(N) \sum_{n=M+1}^N (n/x)^\rho c_n(x) \\ &= -\tfrac{1}{2}\sigma_2(N) \{1 + o(1) - \theta_2(N+1, x) - \theta_3(M, x)\}, \\ S_1 &\leq \sigma_1(M) \sum_{n=1}^M (n/x)^\rho c_n(x) \leq \sigma_2(N) \theta_3(M, x), \end{aligned}$$

since $\sigma_1(M) \leq \sigma_1(N) < \sigma_2(N)$. Combining these estimates for S_1, S_2, S_3 we find from (11) that $x^{-\rho} V_a^k(x) \rightarrow -\infty$ as $H \rightarrow \infty$ contradicting (8). This completes the proof.

Lemma 5. If $s(t) = s_n$ for $n \leq t < n+1$ and $s_n = O(1)$ ($n \rightarrow \infty$), then

$$V_a^k(x) - \int_0^\infty (a/\pi x)^{1/2} \exp\{-ax^{-1}(t-x)^2\} s(t) dt \rightarrow 0 \quad (x \rightarrow \infty).$$

Proof. It suffices to prove that

$$\sum_{n=0}^{\infty} |c_n(x) - \int_n^{n+1} (a/\pi x)^{1/2} \exp\{-ax^{-1}(t-x)^2\} dt| \rightarrow 0 \quad (x \rightarrow \infty). \quad (12)$$

Denoting the summand in (12) by $d_n(x)$, we find from the condition (iii) on $c_n(x)$ with $\sigma = 0$ that, as $x \rightarrow \infty$

$$\begin{aligned} \sum_{|n-x| > x^\delta} d_n(x) &\leq o(1) + \left(\int_0^{x-x^\delta} + \int_{x+x^\delta}^\infty \right) (a/\pi x)^{1/2} \exp \{-ax^{-1}(t-x)^2\} dt \\ &= o(1) + \left(\int_{-x^\delta}^{-x^{\delta-\frac{1}{2}}} + \int_{x^{\delta-\frac{1}{2}}}^\infty \right) (a/\pi)^{1/2} \exp(-au^2) du = o(1); \end{aligned}$$

on the other hand, using Lemma 2(b) with $\sigma = 0$,

$$\sum_{|n-x| \leq x^\delta} d_n(x) \leq \sum_{|n-x| \leq x^\delta} \epsilon \int_n^{n+1} (a/\pi x)^{1/2} \exp \{-ax^{-1}(t-x)^2\} dt < \epsilon,$$

for $x > x_0(\epsilon)$. This proves (12).

Lemma 6. If s_n is a real sequence satisfying

$$\lim_{\epsilon \rightarrow 0+} \liminf_{m \rightarrow \infty} \min_{m \leq n \leq m + \epsilon \sqrt{m}} (s_n - s_m) \geq 0, \quad (13)$$

$$(a/\pi x)^{1/2} \int_0^\infty \exp \{-ax^{-1}(t-x)^2\} s(t) dt \rightarrow s(x \rightarrow \infty),$$

where $s(t) = s_n$ for $n \leq t < n+1$ and if $s_n = O(1)$ ($n \rightarrow \infty$), then $s_n \rightarrow s$ ($n \rightarrow \infty$).

Taking $s = 0$ this lemma is proved exactly like its case $a = \frac{1}{2}$ ([3], pp. 313–314).

Lemma 7. If $s_n \rightarrow s(B_{a,\gamma})$ then, for $r \geq 0$,

$$a^{r+1} \Gamma(r+1) \exp(-ax) \sum_{n=N}^\infty \frac{(ax)^{na+\gamma-1}}{\Gamma(na+\gamma+r)} S_n^r \rightarrow s(x \rightarrow \infty), \quad (14)$$

$$t_r(n) \equiv a^r \frac{\Gamma(r+1) \Gamma(na+\gamma)}{\Gamma(na+\gamma+r)} S_n^r \rightarrow s(B_{a,\gamma}). \quad (15)$$

(14) is known ([1], Lemma 4) and (15) follows readily from (14).

Lemma 8. If $s_n \rightarrow s(B_{a,\gamma})$ and $s_n^r = O(n^\sigma)$ for some $r \geq 0$, $\sigma \geq \frac{1}{2}$, then $s_n^{r+1} = O(n^{\sigma-1/2})$, and, more generally, for an integer q such that $1 \leq q \leq 2\sigma$, $s_n^{r+q} = O(n^{\sigma-q/2})$.

Proof. We have identically, for $m < n$,

$$s_n^{r+1} - s_m^{r+1} = \sum_{\nu=m+1}^n (s_\nu^{r+1} - s_{\nu-1}^{r+1}) = \sum_{\nu=m+1}^n \frac{r+1}{\nu} (s_\nu^r - s_{\nu-1}^r),$$

(see [3], p. 122). By hypothesis, $s_\nu^r = O(\nu^\sigma)$ as $\nu \rightarrow \infty$ and so $s_\nu^{r+1} = O(\nu^\sigma)$. Hence as $m \rightarrow \infty$, we have uniformly for $m \leq n \leq m + \epsilon \sqrt{m}$, $0 < \epsilon < 1$,

$$s_n^{r+1} - s_m^{r+1} = \sum_{\nu=m+1}^n O(\nu^{\sigma-1}) = \epsilon O(m^{\sigma-1/2}). \quad (16)$$

By the definition of $t_r(n)$ in (15) and Stirling's approximation,

$$t_r(n) = s_n^r \left\{ 1 + \frac{1}{n} w_r(n) \right\}, \quad w_r(n) = O(1), \quad (17)$$

so that we can write

$$t_{r+1}(n) - t_{r+1}(m) = s_n^{r+1} - s_m^{r+1} + \frac{1}{n} s_n^{r+1} w_{r+1}(n) - \frac{1}{m} s_m^{r+1} w_{r+1}(m). \quad (18)$$

Using (16) in (18), we get uniformly for $m \leq n \leq m + \epsilon \sqrt{m}$, $0 < \epsilon < 1$, as $m \rightarrow \infty$

$$t_{r+1}(n) - t_{r+1}(m) = \epsilon O(m^{\sigma-1/2}) + O(m^{\sigma-1});$$

$$\lim_{\epsilon \rightarrow 0} \limsup_{m \rightarrow \infty} \max_{m \leq n \leq m + \epsilon \sqrt{m}} |t_{r+1}(n) - t_{r+1}(m)| m^{-\sigma+1/2} = 0.$$

This and $t_r(n) \rightarrow s(B_{a,\gamma})$ which is a consequence of the hypothesis $s_n \rightarrow s(B_{a,\gamma})$ by Lemma 7, together imply $t_r(n) = O(n^{\sigma-1/2})$ on appeal to Lemma 4 with $t_r(n)$, $\sigma - \frac{1}{2}$ instead of s_n , ρ respectively and with the V_a^k transform specialized to the $(B_{a,\gamma})$ transform. Finally, we have also $s_n^{r+1} = O(n^{\sigma-1/2})$ because of (17).

The more general conclusion of the lemma follows from successive repetitions q times of the above argument.

4. Theorems

Theorem 1. Suppose that $s_n = O(n^\sigma)$ ($n \rightarrow \infty$) for some $\sigma \geq 0$. Then $s_n \rightarrow s(V_a^k)$ if and only if

$$\left(\frac{a}{\pi x}\right)^{1/2} \sum_{n=0}^{\infty} \exp[-ax^{-1}(n-x)^2] s_n \rightarrow s(x \rightarrow \infty).$$

The proof is exactly like that of Satz II of Faulhaber [2]. Combining Theorem 1 with Lemma 1 we have

Theorem 2. Suppose that $s_n = O(n^\sigma)$ ($n \rightarrow \infty$) for some $\sigma \geq 0$. Then $s_n \rightarrow s(V_a^k)$ if and only if $s_n \rightarrow s(B_{2a,\gamma})$.

Theorem 3. If s_n is a real sequence satisfying (13) and $s_n \rightarrow s(V_a^k)$ then $s_n \rightarrow s(n \rightarrow \infty)$.

Proof. Since (13) is the special case $\rho = 0$ of (1), it follows from Lemma 4 that $s_n = O(1)$ and thereafter the desired conclusion is obvious from Lemmas 5 and 6.

Theorem 4. If s_n is a real sequence satisfying (1) and $s_n \rightarrow s(V_a^k)$, then $s_n \rightarrow s(C, 2\rho)$.

Proof. After Theorem 3, we need prove only the case $\rho > 0$. Also, $s_n = O(n^\rho)$ by lemma 4, so that we may appeal to Theorem 2 and replace hypothesis $s_n \rightarrow s(V_a^k)$ by $s_n \rightarrow s(B_{2a,\gamma})$.

If p is the greatest integer less than $2\rho + 1$ then

$$0 < \mu = 2\rho - (p - 1) \leq 1$$

and it follows from the more general conclusion of Lemma 8 with $r = 0$ that

$$s_n^{p-1} = O(n^{\rho-(p-1)/2}) = O(n^{\mu/2}).$$

Hence, exactly as elsewhere ([4], proof of Theorem 2, 439-40) we obtain, for $m \leq n \leq m + \epsilon \sqrt{m}$, $0 < \epsilon < 1$, uniformly as $m \rightarrow \infty$

$$s_n^{2\rho} - s_m^{2\rho} = \epsilon^\mu O(1) + \epsilon O(m^{(\mu-1)/2}), \quad s_m^{2\rho} = O(m^{\mu/2}). \quad (19)$$

Further, the definition of $t_r(n)$ in (15) gives, by (17),

$$\begin{aligned} t_{2\rho}(n) - t_{2\rho}(m) &= s_n^{2\rho} - s_m^{2\rho} + \frac{1}{n} s_n^{2\rho} w_{2\rho}(n) - \frac{1}{m} s_m^{2\rho} w_{2\rho}(m) \\ &= \epsilon^\mu O(1) + \epsilon O(m^{(\mu-1)/2}) + O(m^{\mu/2-1}) \end{aligned}$$

on account of (19). Thus

$$\lim_{\epsilon \rightarrow 0+} \limsup_{m \rightarrow \infty} \max_{m \leq n \leq m + \epsilon \sqrt{m}} |t_{2\rho}(n) - t_{2\rho}(m)| = 0,$$

while our assumption $s_n \rightarrow s(B_{2a}, \gamma)$ implies $t_{2\rho}(n) \rightarrow s(B_{2a}, \gamma)$ by Lemma 7. Now appealing to Theorem 3, with s_n replaced by $t_{2\rho}(n)$ and summability (V_a^k) by summability (B_{2a}, γ) , we see that

$$t_{2\rho}(n) \rightarrow s \quad \text{whence} \quad s_n^{2\rho} \rightarrow s$$

as required, by (17).

5. Some standard cases of summability method (V_a^k)

In addition to the general Borel method (B_a, γ) and Borel method which is $(B_1, 1)$, the following are special cases of the method (V_a^k) .

I. The Valiron method (V_a) which corresponds to the transforms of s_n given by

$$V_a(x) = \left(\frac{a}{\pi x}\right)^{1/2} \sum_{n=0}^{\infty} \exp[-ax^{-1}(n-x)^2] s_n, \quad x > 0,$$

is a special case. For, $V_a(x)$ is $V_a^k(x)$ of § 2, with $g_k + R_k \equiv 0$ in condition (ii) imposed on $c_n(x)$, and with condition (iii) on $c_n(x)$ satisfied in consequence of a result stated by Faulhaber (Hilfssatz 1 with $p = 0$).

II. The Euler method (E_p) , $p > 0$, the Hardy-Littlewood method (T_a) , $0 < a < 1$, and the method (S_β) , $0 < \beta < 1$, due to Meyer-König and Vermes, correspond to transforms of s_n which may be written

$$\sum_{n=0}^{\infty} c_{n,m} s_n, \quad m = 1, 2, 3, \dots,$$

$$\text{where for } (E_p) : c_{n,m} = 2^{-pm} \binom{m}{n} (2^n - 1)^{m-n},$$

$$\text{for } (T_a) : c_{n,m} = (1-a)^{m+1} \binom{n}{m} a^{n-m},$$

$$\text{for } (S_\beta) : c_{n,m} = (1-\beta)^{m+1} \binom{m+n}{n} \beta^n,$$

with the convention that $\binom{v}{r} = 0$ for $v < r$. In all these cases we may define

$$c_n(x) = \begin{cases} c_{n,m} & \text{for } x_m \leq x < x_{m+1} \\ 0 & \text{for } 0 < x < x_1, \end{cases} \quad (20)$$

where $x_m = \lambda m$ with a suitable constant λ in each case. We can verify, using arguments of Faulhaber (proof of Hilfssatz 5) with trivial modifications, that the $c_n(x)$ of (20) satisfy for $x = x_m$ the conditions required of the $c_n(x)$ in the definition of $V_a^k(x)$ in § 2, provided we choose

$$\text{for } (E_p) : a = 2^p / (2^{p+1} - 2), \quad \lambda = 2^{-p},$$

$$\text{for } (T_a) : a = (1 - \alpha) / 2\alpha, \quad \lambda = (1 - \alpha)^{-1},$$

$$\text{for } (S_\beta) : a = (1 - \beta) / 2, \quad \lambda = \beta(1 - \beta)^{-1}.$$

Finally we can verify that, if the $c_n(x)$ of (20) satisfy the required conditions for $x = x_m$, where x_m is any increasing unbounded sequence with $x_{m+1} - x_m = O(1)$ then they satisfy the conditions for all $x > 0$. In particular, taking $x_m = \lambda m$ we see that the transforms E_p , T_a , S_β are all special cases of the V_a^k transform.

Acknowledgements

The author is indebted to the late Prof. C T Rajagopal for his help in the preparation of this paper, especially in the proofs of Lemma 8 and Theorem 4.

References

- [1] Borwein D 1969 *Can. J. Math.* **21** 740
- [2] Faulhaber G 1956 *Math. Z.* **66** 34
- [3] Hardy G H 1949 *Divergent series* (Oxford : Clarendon Press)
- [4] Rajagopal C T 1960 *J. Indian Math. Soc. (N.S.)* **24** 433
- [5] Sitaraman Y 1967 *Math. Z.* **95** 34

On the mean square value of Hurwitz zeta function

M J NARLIKAR

701, Colaba Housing Colony, Homi Bhabha Road, Bombay 400 005, India

MS received 9 August 1980 ; revised 6 October 1980

Abstract. R Balasubramanian has shown that

$$\int_1^T |\zeta(\tfrac{1}{2} + it)|^2 dt = T \log \frac{T}{2\pi} + (2\gamma - 1) T + O(T^{\theta+\epsilon})$$

with $\theta = \frac{1}{3}$. In this paper we develop a hybrid analogue for the mean square value of the Hurwitz zeta function $\zeta(s, a)$ and show that (i) new asymptotic terms arise in the expression for $\zeta(s, a)$ which are not present in the above expression for the ordinary zeta function and (ii) the corresponding error term is given by

$$O(T^{5/12} \log^2 T) + O\left(\frac{\log T}{||2a||}\right)$$

for $0 < a < 1$.

Keywords. Hybrid analogue for L-series; Balasubramanian's theorem.

1. Introduction

In a recent paper, Balasubramanian [1] proved the following theorem.

Theorem 1 :

$$\int_0^T |\zeta(\tfrac{1}{2} + it)|^2 dt = T \log (T/2\pi) + (2\gamma - 1) T + O(T^{\theta+\epsilon}),$$

with $\theta = \frac{1}{3}$.

The above theorem with $\theta = 5/12$ is a deep theorem of Titchmarsh [7] and Balasubramanian was the first to make a major improvement on it. Later Good [2] gave a different proof of part of Balasubramanian's result. Heath-Brown [3], by a different set of more powerful ideas, proved Theorem 1 and a very important result on the fourth power mean.

In an attempt to get a hybrid analogue, we are led to the consideration of the mean square of the Hurwitz zeta function on the critical line. So we try to get a result similar to Theorem 1, for Hurwitz zeta function $\zeta(s, a)$, defined by

$$\zeta(s, a) = \sum_{n=0}^{\infty} \left(\frac{1}{(n+a)^s} - \int_n^{n+1} \frac{du}{(u+a)^s} \right) - \frac{a^{1-s}}{1-s},$$

for $0 < a < 1$ and $\text{Re } s > 0$.

Using the approximate functional equation for $\zeta(s, a)$, Rane [5] has already proved

Theorem 2 :

$$\int_1^T |\zeta(\tfrac{1}{2} + it, a)|^2 dt = A(a, T) + O(a^{-1/2} T^{1/2} \log T)$$

where $A(a, T) = T \log(T/2\pi) + T[C(a) + \gamma - 1] - 1/a$

$$\text{and } C(a) = \lim_{m \rightarrow \infty} \left[\sum_{n=0}^m \frac{1}{n+a} - \log m \right].$$

We shall use a much more delicate analysis introduced by Titchmarsh in [7] and prove

Theorem 3 : If $0 < a < 1$ and $A(a, T)$ is as in Theorem 2, then for $a \neq \frac{1}{2}$

$$\begin{aligned} \int_0^T |\zeta(\tfrac{1}{2} + it, a)|^2 dt \\ = A(a, T) + A_1(a) T^{1/2} + O(T^{5/12} \log^2 T) + O\left(\frac{\log T}{\|a\|}\right), \end{aligned}$$

$$\begin{aligned} \text{where } A_1(a) = 2\pi \left[-2H(a) + \int_{1-a}^1 \frac{\cos \pi(u^2 - u - \tfrac{1}{4})}{\cos \pi u} du \right. \\ \left. + 2 \sum_{r=1}^{\infty} \int_r^{\infty} \cos \pi(u^2 - \tfrac{1}{4}) du \right] < -(\sqrt{2\pi}) H(a). \end{aligned}$$

$$\text{with } H(a) = \int_{(1-a)/2}^{(1+a)/2} G^2(u) du + \int_{a/2}^{1-a/2} G^2(u) du$$

$$\text{and } G(u) = \cos 2\pi(u^2 - u - \tfrac{1}{8}) / \cos 2\pi u,$$

and for a real number a ,

$\|a\|$ = distance between a and the integer nearest to a .

For $a = \frac{1}{2}$,

$$\begin{aligned} \int_1^T |\zeta(\tfrac{1}{2} + it, \tfrac{1}{2})|^2 dt = A(\tfrac{1}{2}, T) + A_2 T^{1/2} \log T + A_3 T^{1/2} \\ + O(T^{5/12} \log^2 T), \end{aligned}$$

$$\text{where } A_2 = -1/4 \sqrt{2\pi},$$

$$\text{and } A_3 = \frac{1}{\sqrt{2\pi}} \left[\tfrac{1}{4} \log 2\pi - 4 \log 2 - 2\gamma - 4\pi H(\tfrac{1}{2}) + \tfrac{1}{2} + A_0 \right],$$

$$\text{where } A_0 = 8\sqrt{2} \sum_{r=1}^{\infty} \int_r^{\infty} \frac{\sin \pi(u^2 + u + \tfrac{1}{4})}{(2u+1)^2} du.$$

Remarks : (1) It is surprising that the term $A_1(a)$ is different from zero for $0 < a < 1$, $a \neq \frac{1}{2}$ and $A_2 \neq 0$ for $a = \frac{1}{2}$.

(2) In this paper, we apply mainly the methods of Titchmarsh [7]. It is obvious from the proof of Lemma 20 in this paper why Balasubramanian's [1] multiple average integral method is not applicable in our case.

(3) The results of this paper, as well as those of Rane [5] have been announced by Ramachandra in [4].

2. Some preliminary results and a few definitions

Lemma P1 :

$$\begin{aligned} & \int_{2\pi a^2}^{\infty} \exp \left[i \left(t \log \frac{t}{2\pi ab} - t - \frac{\pi}{4} \right) \right] dt \\ &= 2\pi a \exp \left[2\pi i a^2 \left(\log \frac{a}{b} - \frac{1}{2} \log^2 \frac{a}{b} \right) \right] \\ & \quad \times \int_{a \log a/b}^{\infty} \exp \left[i\pi \left(\theta^2 - \frac{1}{4} \right) \right] d\theta + O \left(\frac{1}{(a-b)^4} \right), \end{aligned}$$

for $0 < b < a < Ab$, A, a, b being real numbers.

Titchmarsh proves the above lemma in [7] with a and b as positive integers. It is not difficult to see that practically the same proof works in our case.

Lemma P2 : For $0 \leq r$ and $0 < a \leq 1$, and r an integer

$$\begin{aligned} & \int_{r+a}^{\infty} \exp [i\pi (\theta^2 - \tfrac{1}{4})] d\theta \\ &= - \frac{(-1)^r \exp [i\pi (a^2 + 2ra - \tfrac{1}{4})]}{2\pi i (r+a)} + \frac{(-1)^r \exp [i\pi (a^2 + 2ra - \tfrac{1}{4})]}{4\pi^2 (r+a)^3} \\ & \quad + O \left(\frac{1}{(r+a)^4} \right). \end{aligned}$$

Proof : Follows from integration by parts.

Lemma P3 : If $0 < a \leq 1$ and $0 \leq \sigma \leq 1$, $\sigma = \operatorname{Re} s$,

$$k_1 = \left[\sqrt{\frac{t}{2\pi}} - a \right], \quad k = \left[\sqrt{\frac{t}{2\pi}} \right], \quad \beta = k - k_1, \quad s = \sigma + it, \text{ then}$$

$$\zeta(\sigma + it, a)$$

$$\begin{aligned} &= \sum_{0 \leq n \leq k_1} \frac{1}{(n+a)^s} + \left(\frac{2\pi}{t} \right)^{it} \exp \left[i \left(t + \frac{\pi}{4} \right) \right] \\ & \quad \times \sum_{1 \leq m \leq k} \left[\frac{\exp(-2\pi i m a)}{m^{\frac{1}{2}-it}} \right] + g_3(t) + O(t^{-3/4}), \end{aligned} \quad (1)$$

$$\text{where } g_3(t) = (2\pi/t)^{1/4} i^{\beta} (-1)^k \exp \left[-\frac{7}{8} \pi i + \frac{\pi i a^2}{2} + i \frac{t}{2} \log \frac{2\pi}{t} \right.$$

$$\left. + \pi i (2\beta\gamma - 2a\gamma - a\beta) \right] \psi(t),$$

with

$$\psi(t) = \frac{\cos \left[t - \sqrt{2\pi t} (2k+1 + \alpha - \beta) + \pi \left\{ \frac{(\alpha - \beta)^2}{2} + (\alpha - \beta)(2k+1) - \frac{1}{8} \right\} \right]}{\cos [\sqrt{2\pi t} - (\alpha - \beta)\pi]}$$

and $y = (t/2\pi)^{1/2}$.

Proof. This is the approximate functional equation for $\zeta(s, \alpha)$, and as in Rane's work [5] it is obtained in the same way as the approximate functional equation for $\zeta(s)$ in Titchmarsh [6], equation (4.17.1).

Remark : If we denote the four terms on the R.H.S. of equation (1) by $g_i(\alpha, t)$, $i = 1, 2, 3, 4$, then it is easy to see that $g_3(\alpha, t) = O(t^{-1/4})$.

For the proofs of the following two lemmas, we refer to Titchmarsh [7].

Lemma P4 : If $f(x)$ is real and in C^2 over $a \leq x \leq b$, and $f''(x) \geq r > 0$ (or $\leq -r < 0$) in $a < x < b$ and $|f'(b) - f'(a)| = (b-a)R$, then if n denotes integer,

$$\sum_{a \leq n \leq b} \exp [2\pi i f(n)] = O \left\{ \frac{(b-a)R}{\sqrt{r}} \right\} + O(1/\sqrt{r}).$$

Lemma P5 : If $f(x)$ is real and is in C^3 over $a \leq x \leq b$, and $f'''(x) \geq r_2 > 0$ (or $\leq -r_2 < 0$) in $a < x < b$, then if $|f''(b) - f''(a)| = (b-a)R_2$,

$$\sum_{a \leq n \leq b} \exp [2\pi i f(n)] = O \left\{ \frac{(b-a)R_2^{1/3}}{r_2^{1/6}} \right\} + O \left\{ \frac{(b-a)^{1/2} R_2^{1/6}}{r_2^{1/3}} \right\}.$$

We now give the notation to be used in the proof :

$$K = [(T/2\pi)^{1/2}], \quad K_1 = [(T/2\pi)^{1/2} - \alpha],$$

where for a real number α , $[a]$ = the largest integer not greater than a .

Q_i , $i = 1, 2, 3, 4$, will be integers with $Q_i = K + O(1)$, $Q_i \leq K$.

$$\begin{aligned} B_1 = 2 \operatorname{Re} \sum_{n=0}^{K_1} \frac{1}{(n+\alpha)^{1/2}} \sum_{n+1 \leq q}^{Q_1} (-1)^{q-1} \int_{2\pi q^2}^{2\pi(q+\alpha)^2} i \left(\frac{2\pi}{t} \right)^{1/4} \\ \times \exp \left[it \log(n+\alpha) + \frac{it}{2} \log \frac{2\pi}{t} + \frac{it}{2} + i\pi(2y - 2\alpha y - \alpha) \right. \\ \left. + \frac{i\pi}{8} + \frac{i\pi\alpha^2}{2} \right] \psi(t) dt, \end{aligned}$$

$$\begin{aligned} B_2 = 2 \operatorname{Re} \sum_{n=0}^{K_1} \frac{1}{(n+\alpha)^{1/2}} \sum_{n \leq q}^{Q_2} (-1)^q \int_{2\pi(q+\alpha)^2}^{2\pi(q+1)^2} \left(\frac{2\pi}{t} \right)^{1/4} \\ \times \exp \left[it \log(n+\alpha) + \frac{it}{2} \log \frac{2\pi}{t} + \frac{it}{2} + i\pi(-2\alpha y) \right] \end{aligned}$$

$$+ \frac{i\pi}{8} + \frac{i\pi\alpha^2}{2} \Big] \psi(t) dt,$$

$$B_3 = 2 \operatorname{Re} \sum_{1 \leq n}^K \frac{1}{n^{1/2}} \sum_{n \leq q}^{Q_3} (-1)^a \int_{2\pi q^2}^{2\pi(q+\alpha)^2} \psi(t) \cdot \left(\frac{2\pi}{t}\right)^{1/4} \\ \times \exp \left[it \log \left(\frac{n}{y} \right) + \frac{it}{2} - 2\pi i n \alpha - \frac{i\pi\alpha^2}{2} + \frac{9}{8} \pi i \right. \\ \left. - i\pi(2y - 2\alpha y - \alpha) \right] dt,$$

$$B_4 = 2 \operatorname{Re} \sum_{1 \leq n}^K \frac{1}{n^{1/2}} \sum_{n \leq q}^{Q_4} (-1)^a \int_{2\pi(q+\alpha)^2}^{2\pi(q+1)^2} \psi(t) \cdot \left(\frac{2\pi}{t}\right)^{1/4} \\ \times \exp \left[it \log \left(\frac{n}{y} \right) + \frac{it}{2} - 2\pi i n \alpha - \frac{i\pi\alpha^2}{2} + \frac{9}{8} \pi i + 2\pi i \alpha y \right] dt,$$

where $\psi(t)$ is as defined in Lemma P3.

$$C_1 = 2 \operatorname{Re} \sum_{n=0}^{K_1} \frac{1}{(n+\alpha)^{1/2}} \sum_{n+n^{1/4} \leq q}^{Q_1} (-1)^{a-1} \int_{2\pi q^2}^{2\pi(q+\alpha)^2} i \psi(t) \cdot \left(\frac{2\pi}{t}\right)^{1/4} \\ \times \exp \left[it \log \left(\frac{n+\alpha}{y} \right) + \frac{it}{2} + \frac{i\pi}{8} + \frac{i\pi\alpha^2}{2} + i\pi(2y - 2\alpha y - \alpha) \right] dt,$$

$$C_2 = 2 \operatorname{Re} \sum_{n=0}^{K_1} \frac{1}{(n+\alpha)^{1/2}} \sum_{n+n^{1/4} \leq q}^{Q_2} (-1)^a \int_{2\pi(q+\alpha)^2}^{2\pi(q+1)^2} \psi(t) \left(\frac{2\pi}{t}\right)^{1/4} \\ \exp \left[it \log \left(\frac{n+\alpha}{y} \right) + \frac{it}{2} + \frac{i\pi}{8} + \frac{i\pi\alpha^2}{2} - 2\pi i \alpha y \right] dt,$$

$$C_3 = 2 \operatorname{Re} \sum_{n=1}^K \frac{1}{n^{1/2}} \sum_{n+n^{1/4} \leq q}^{Q_3} (-1)^a \int_{2\pi q^2}^{2\pi(q+\alpha)^2} \psi(t) \cdot \left(\frac{2\pi}{t}\right)^{1/4} \\ \times \exp \left[it \log \left(\frac{n}{y} \right) + \frac{it}{2} - 2\pi i n \alpha - \frac{i\pi\alpha^2}{2} \right. \\ \left. + \frac{9}{8} \pi i - i\pi(2y - 2\alpha y - \alpha) \right] dt,$$

$$C_4 = 2 \operatorname{Re} \sum_{n=1}^K \frac{1}{n^{1/2}} \sum_{n+n^{1/4} \leq q}^{Q_4} (-1)^a \int_{2\pi(q+\alpha)^2}^{2\pi(q+1)^2} \psi(t) \cdot \left(\frac{2\pi}{t}\right)^{1/4} \\ \times \exp \left[it \log \left(\frac{n}{y} \right) + \frac{it}{2} - 2\pi i n \alpha - \frac{i\pi\alpha^2}{2} + \frac{9}{8} \pi i + 2\pi i \alpha y \right] dt,$$

$$B_5 = 2 \operatorname{Re} \sum_{\substack{m \leq n \leq m+n^{1/4} \\ n \leq K_1}} \frac{\exp(2\pi i m a)}{m^{1/2} (n+a)^{1/2}} \\ \times \int_{2\pi(n+a)^2}^{\infty} \exp[i\phi(t) - it \log \{m(n+a)\}] dt$$

where $\phi(t) = t \log(t/2\pi) - t - \pi/4$.

$$B_6 = 2 \operatorname{Re} \sum_{\substack{n < m \leq n+n^{1/4} \\ m \leq K}} \frac{\exp(2\pi i m a)}{m^{1/2} (n+a)^{1/2}} \\ \times \int_{2\pi m^2}^{\infty} \exp[i\phi(t) - it \log \{m(n+a)\}] dt,$$

$$C_5 = 2 \operatorname{Re} \sum_{m \leq n \leq K_1} \frac{\exp(2\pi i m a)}{m^{1/2} (n+a)^{1/2}} \\ \times \int_T^{\infty} \exp[i\phi(t) - it \log \{m(n+a)\}] dt,$$

$$C_6 = 2 \operatorname{Re} \sum_{n < m \leq K} \frac{\exp(2\pi i m a)}{m^{1/2} (n+a)^{1/2}} \\ \times \int_T^{\infty} \exp[i\phi(t) - it \log \{m(n+a)\}] dt,$$

$$R(B_5) = 2 \operatorname{Re} \sum_{\substack{m+n^{1/4} < n < K_1}} \frac{\exp(2\pi i m a)}{m^{1/2} (n+a)^{1/2}} \\ \times \int_{2\pi(n+a)^2}^{\infty} \exp[i\phi(t) - it \log \{m(n+a)\}] dt,$$

$$R(B_6) = 2 \operatorname{Re} \sum_{\substack{n+n^{1/4} < m \leq K}} \frac{\exp(2\pi i m a)}{m^{1/2} (n+a)^{1/2}} \\ \times \int_{2\pi m^2}^{\infty} \exp[i\phi(t) - it \log \{m(n+a)\}] dt.$$

From Lemma P3 and the remark after it, we get

Lemma P6 :

$$\int_1^T |\zeta(\tfrac{1}{2} + it, a)|^2 dt = I_1 + I_2 + I_3 + I_4 + I_5 + O(T^\epsilon)$$

where g_i are defined as in the remark after Lemma P3 and

$$I_1 = \int_1^T |g_1(a, t) + g_2(a, t)|^2 dt,$$

$$I_2 = \int_1^T |g_3(a, t)|^2 dt,$$

$$I_3 = \int_1^T \{(g_1 \bar{g}_3 + g_3 \bar{g}_1)\}(a, t) dt,$$

$$I_4 = \int_1^T \{(g_2 \bar{g} + g_3 \bar{g}_2)\}(a, t) dt,$$

and
$$I_5 = \int_1^T \{\bar{g}_1 g_4 + g_1 \bar{g}_4 + \bar{g}_2 g_4 + g_2 \bar{g}_4\}(a, t) dt.$$

3. Proof of theorem 3

We shall estimate I_j , $1 \leq j \leq 5$ one by one and first prove Theorem 3'.

Theorem 3' : With the notation of Theorem 3 if $\alpha \neq \frac{1}{2}$,

$$\begin{aligned} \int_1^T |\zeta(\tfrac{1}{2} + it, \alpha)|^2 dt &= A(\alpha, T) + A_1(\alpha) T^{1/2} + 2V_3 + 2V_5 - 2V_{4,2} - 2V_{6,2} \\ &\quad - C_5 - C_6 + C_1 + C_2 + C_3 + C_4 + R(B_5) + R(B_6) \\ &\quad + O(T^{3/8}) + O\left(\frac{\log T}{\|2\alpha\|}\right), \end{aligned}$$

and for $\alpha = \frac{1}{2}$

$$\begin{aligned} \int_1^T |\zeta(\tfrac{1}{2} + it, \tfrac{1}{2})|^2 dt &= A(\tfrac{1}{2}, T) + A_2 T^{1/2} \log T + A_3 T^{1/2} + 2V_3 + 2V_5 \\ &\quad - 2V_{4,2} - 2V_{6,2} - C_5 - C_6 + R(B_5) + R(B_6) \\ &\quad + C_1 + C_2 + C_3 + C_4 + O(T^{3/8}), \end{aligned}$$

where

$$V_3 = \sum_{1 \leq m < n \leq K} \frac{\sin \left[T \log \left(\frac{n}{m} \right) + 2\pi(m-n)\alpha \right]}{m^{1/2} n^{1/2} \log \left(\frac{n}{m} \right)},$$

$$V_5 = \sum_{0 \leq m < n \leq K_1} \frac{\sin \left[T \log \left(\frac{n+\alpha}{m+\alpha} \right) \right]}{(m+\alpha)^{1/2} (n+\alpha)^{1/2} \log \left(\frac{n+\alpha}{m+\alpha} \right)},$$

$$V_{4,2} = \sum_{m+m^{1/4} < n \leq K} \frac{\sin \left[2\pi n^2 \log \left(\frac{n}{m} \right) + 2\pi(m-n)\alpha \right]}{m^{1/2} n^{1/2} \log \left(\frac{n}{m} \right)},$$

$$V_{6,2} = \sum_{m+m^{1/4} < n \leq K_1} \frac{\sin \left[2\pi (n+a)^2 \log \left(\frac{n+a}{m+a} \right) \right]}{(m+a)^{1/2} (n+a)^{1/2} \log \left(\frac{n+a}{m+a} \right)}.$$

Then we shall deduce Theorem 3 from Theorem 3'.

Lemma 1: $I_3 = B_1 + B_2 + O(1)$.

Proof: We divide the interval $[2\pi(n+a)^2, T]$ into intervals of the type $[q, q+a]$ and $[q+a, q+1]$ where q is a positive integer and then substitute the value of β in each case.

Lemma 2:

$$B_1 = -4\pi K \int_{(1-a)/2}^{(1+a)/2} G^2(u) du + C_1 + O(T^{3/8}).$$

Proof: Of the summation over q in B_1 , we make two parts as $n < q < n + n^{1/4}$ and $n + n^{1/4} \leq q \leq Q_1$. The second part immediately gives C_1 . If the first part is defined to be $S(B_1)$,

$$\begin{aligned} S(B_1) &= 2 \operatorname{Re} \sum_{n \geq 0} \frac{1}{(n+a)^{1/2}} \sum_{n < q < n+n^{1/4}} (-1)^{q-1} \int_{2\pi q^2}^{2\pi(q+a)^2} i\psi(t) \left(\frac{2\pi}{t} \right)^{1/4} \\ &\quad \times \exp \left[it \log \left(\frac{n+a}{y} \right) + \frac{it}{2} + \frac{i\pi}{8} + \frac{i\pi a^2}{2} \right. \\ &\quad \left. + i\pi(2y - 2ay - a) \right] dt. \end{aligned}$$

Let $q = n + r$, $1 \leq r < n^{1/4}$ and $(2\pi t)^{1/2} - (a-1)\pi = 2\pi(q+u)$, $r+u+(a-1)/2 = \omega$. Then clearly $\psi(t) = G(u)$ and

$$\begin{aligned} S(B_1) &= 8\pi \operatorname{Re} \sum_{n \geq 0} \frac{1}{(n+a)^{1/2}} \sum_{n < q < n+n^{1/4}} (-1)^{n+r-1} \int_{(1-a)/2}^{(1+a)/2} G(u) (n+\omega)^{1/2} \\ &\quad \times \exp \left[2\pi i(1-a)(n+\omega) + \frac{5i\pi}{8} - i\pi a + 2\pi i(n+\omega)^2 \right. \\ &\quad \left. \times \log \left(\frac{n+a}{n+\omega} \right) + i\pi(n+\omega)^2 + \frac{i\pi a^2}{2} \right] du. \end{aligned}$$

$$\begin{aligned} \text{Let } \cos \left[2\pi(n+\omega)(1-a) + \frac{5\pi}{8} - \pi a + 2\pi(n+\omega)^2 \log \left(\frac{n+a}{n+\omega} \right) \right. \\ \left. + \pi(n+\omega)^2 + \frac{\pi a^2}{2} \right] = \cos X. \end{aligned}$$

$$\text{Then } \cos X = \cos(-X) = \cos \left[-\pi n^2 + (2\omega^2 - 2\omega a + a^2 - 2\omega)\pi \right]$$

$$+ \pi\alpha - \frac{\pi\alpha^2}{2} - \frac{5\pi}{8} + \frac{2(\omega - \alpha)^3}{3(n + \alpha)} - \pi + O\left(\frac{r^4}{(n + \alpha)^2}\right)].$$

Now substituting the values of ω we get

$$\cos X = (-1)^n \cos \pi \left[2u^2 + 4u(r - 1) + \frac{7}{8} + \frac{2\left(u + r - \frac{1 + \alpha}{2}\right)^3}{(n + \alpha)} \right. \\ \left. + O\left(\frac{r^4}{n^2}\right) \right].$$

$$\text{Let } \pi \left[2u^2 + 4u(r - 1) + \frac{7}{8} \right] = A, \quad \frac{2\pi\left(u + r - \frac{1 + \alpha}{2}\right)^3}{3(n + \alpha)} + O\left(\frac{r^4}{n^2}\right) = B,$$

and use the formula $\cos(A + B) = \cos A \cos B - \sin A \sin B$ and note that the substitution of 1 for $\cos B$ and

$$\frac{2\pi\left(u + r - \frac{1 + \alpha}{2}\right)^3}{n + \alpha} \text{ for } \sin B \text{ gives an error } O(T^{3/8}).$$

$$\text{Thus } S(B_1) = 8\pi \sum_n \sum_{r < n^{1/4}} \frac{(-1)^{r-1}}{(n + \alpha)^{1/2}} \int_{(1-\alpha)/2}^{(1+\alpha)/2} G(u) \left(n + r + u + \frac{\alpha - 1}{2}\right)^{1/2} \\ \times \cos 2\pi \left[u^2 + 2u(r - 1) + \frac{7}{16} \right] du - 8\pi \sum_n \sum_{r < n^{1/4}} \frac{(-1)^{r-1}}{(n + \alpha)^{1/2}} \\ \times \int_{(1-\alpha)/2}^{(1+\alpha)/2} G(u) \left(n + r + u + \frac{\alpha - 1}{2}\right)^{1/2} \sin 2\pi \left[u^2 + 2u(r - 1) \right. \\ \left. + \frac{7}{16} \right] \times \frac{2\pi\left(u + r - \frac{1 + \alpha}{2}\right)^3}{n + \alpha} du + O(T^{3/8}) \\ = S_1 + S_2 + O(T^{3/8}), \text{ say.}$$

Integrating by parts, S_2 is seen to be $O(T^{3/8})$. In S_1 , we replace $(r + u - (1 - \alpha)/2)^{1/2}$ by $(n + \alpha)^{1/2}$ and then change the variable u to $1 - u$.

Then adding both expressions for S_1 and dividing by 2, we get

$$S(B_1) = 8\pi \sum_n \sum_r \frac{(-1)^{r-1}}{(1-\alpha)/2} \int_{(1-\alpha)/2}^{(1+\alpha)/2} G(u) \cos 2\pi \left(u^2 - u + \frac{7}{16} \right) \\ \times \cos 2\pi(2ur - u) du + O(T^{3/8})$$

$$\begin{aligned}
&= 8\pi \sum_n \int_{(1-\alpha)/2}^{(1+\alpha)/2} G(u) \cos 2\pi \left(u^2 - u + \frac{7}{16} \right) \sum_{1 \leq r < n^{1/4}} (-1)^{r-1} \\
&\quad \times \cos 2\pi u (2(r-1) + 1) du + O(T^{3/8}) \\
&= -4\pi \sum_n \int_{(1-\alpha)/2}^{(1+\alpha)/2} G(u) \cos 2\pi \left(u^2 - u - \frac{1}{16} \right) \\
&\quad \times \left[\frac{1 - (-1)^R \cos 4\pi Ru}{\cos 2\pi u} \right] du + O(T^{3/8})
\end{aligned}$$

where $R = [n^{1/4}]$.

Integration by parts of the second term inside the integral sign gives the proof of the Lemma.

Lemma 3 :

$$B_2 = -4\pi K \int_{\alpha/2}^{1-\alpha/2} G^2(u) du + C_2 + O(T^{3/8}).$$

Proof : Similar to that of Lemma 2.

Remark : It is interesting to compare the work of Balasubramanian [1] to see how, for the case $\alpha = 1$, we get finer error terms.

Lemma 4 :

$$I_4 = B_3 + B_4 + O(1).$$

Proof : Similar to that of Lemma 1.

Lemma 5 :

$$B_3 = -4\pi K \int_{(1-\alpha)/2}^{(1+\alpha)/2} G^2(u) du + C_3 + O(T^{3/8}).$$

Proof : Similar to that of Lemma 2.

Lemma 6 :

$$B_4 = -4\pi K \int_{\alpha/2}^{1-\alpha/2} G^2(u) du + C_4 + O(T^{3/8}).$$

Proof : Similar to that of Lemma 2.

Lemma 7 :

$$I_2 = 4\pi K \left[\int_{(1-\alpha)/2}^{(1+\alpha)/2} G^2(u) du + \int_{\alpha/2}^{1-\alpha/2} G^2(u) du \right] + O(T^{3/8}).$$

Proof: We note that

$$I_2 = \sum_{1 \leq Q \leq Q} \left[\int_{2\pi Q^2}^{2\pi(q+\alpha)^2} \psi^2(t) \left(\frac{2\pi}{t}\right)^{1/2} dt + \int_{2\pi(q+\alpha)^2}^{2\pi(q+1)^2} \psi^2(t) \left(\frac{2\pi}{t}\right)^{1/2} dt \right] + O(1)$$

where $Q = K - 1$.

Hence $I_2 = \tilde{S}_1 + \tilde{S}_2 + O(1)$, where $\tilde{S}_1$ and $\tilde{S}_2$ are defined in an obvious way.

For $\tilde{S}_1$ we use the substitution $\sqrt{2\pi t} - (a-1)\pi = 2\pi(q+u)$ and proceed as in Lemma 2 to see that

$$\tilde{S}_1 = 4\pi Q \int_{(1-\alpha)/2}^{(1+\alpha)/2} G^2(u) du + O(T^{3/8}).$$

We let $\sqrt{2\pi t} - a\pi = 2\pi(q+u)$ in $\tilde{S}_2$ and get

$$\tilde{S}_2 = 4\pi Q \int_{\alpha/2}^{1-\alpha/2} G^2(u) du + O(T^{3/8}).$$

Lemma 8:

$$I_5 = O(T^{1/4} \log T).$$

Proof: We assume that

$$\int_0^T |g_1 + g_2|^2(\alpha, t) dt = O(T \log T),$$

which we will prove later.

$$I_5 \leq 2 \left| \int_1^T \{(g_1 + g_2) \bar{g}_4\}(\alpha, t) dt \right|,$$

$$\begin{aligned} \text{and } \int_{T/2}^T \{(g_1 + g_2) \bar{g}_4\}(\alpha, t) dt &\leq \left\{ \int_{T/2}^T |(g_1 + g_2)(\alpha, t)|^2 dt \right. \\ &\quad \times \left. \int_{T/2}^T O(t^{-3/2}) dt \right\}^{1/2} = O(T^{1/4} \log T). \end{aligned}$$

Replacing T by $T/2, T/4, \dots, T/2^n \dots$ and adding, we get the Lemma.

Lemma 9:

$$I_1 = V_1 + V_2 + 2V_3 - 2V_4 + 2V_5 - 2V_6 + V_7 + O(1)$$

$$\text{where } V_1 = \sum_{n=1}^{K_1} \frac{T - 2\pi(n+\alpha)^2}{n+\alpha} + \frac{T-1}{\alpha}, \quad V_2 = \sum_{m=1}^K \frac{T - 2\pi m^2}{m},$$

$$V_3 = \sum_{1 \leq m < n \leq K} \frac{\sin \left[T \log \left(\frac{n}{m} \right) + 2\pi(m-n)\alpha \right]}{m^{1/2} n^{1/2} \log \left(\frac{n}{m} \right)},$$

$$V_4 = \sum_{1 \leq m < n \leq K} \frac{\sin \left[2\pi n^2 \log \left(\frac{n}{m} \right) + 2\pi (m-n) \alpha \right]}{m^{1/2} n^{1/2} \log \left(\frac{n}{m} \right)},$$

$$V_5 = \sum_{0 \leq m < n \leq K_1} \frac{\sin \left[T \log \left(\frac{n+\alpha}{m+\alpha} \right) \right]}{(m+\alpha)^{1/2} (n+\alpha)^{1/2} \log \left(\frac{n+\alpha}{m+\alpha} \right)},$$

$$V_6 = \sum_{0 \leq m < n \leq K_1} \frac{\sin \left[2\pi (n+\alpha)^2 \log \left(\frac{n+\alpha}{m+\alpha} \right) \right]}{(m+\alpha)^{1/2} (n+\alpha)^{1/2} \log \left(\frac{n+\alpha}{m+\alpha} \right)},$$

and $V_7 = V_{7,1} + V_{7,2}$;

$$\text{where } V_{7,1} = 2 \operatorname{Re} \sum_{1 \leq m < n \leq K_1} \frac{\exp(2\pi i m \alpha)}{m^{1/2} (n+\alpha)^{1/2}} \int_{2\pi(n+\alpha)^2}^T \exp i[\phi(t) - t \log \{m(n+\alpha)\}] dt,$$

$$\text{and } V_{7,2} = 2 \operatorname{Re} \sum_{0 \leq m < n \leq K} \frac{\exp(2\pi i m \alpha)}{m^{1/2} (n+\alpha)^{1/2}} \int_{2\pi m^2}^T \exp i[\phi(t) - t \log \{m(n+\alpha)\}] dt.$$

Proof: Follows from the definitions of $g_1(\alpha, t)$ and $g_2(\alpha, t)$. We get

$$\int_1^T |g_2(\alpha, t)|^2 dt = V_2 + 2V_3 - 2V_4 + O(1),$$

$$\int_1^T |g_1(\alpha, t)|^2 dt = V_1 + 2V_5 - 2V_6 + O(1)$$

$$\text{and } \int_1^T (g_1 \bar{g}_2 + \bar{g}_1 g_2)(\alpha, t) dt = V_7 + O(1).$$

Lemma 10: $V_{7,1} = B_5 + R(B_5) - C_5$, where B_5 , $R(B_5)$ and C_5 are as defined in § 2.

Proof: Easy. We only write

$$\text{as } \left[\int_{2\pi(n+\alpha)^2}^T - \int_T^\infty \right],$$

and then for the first part the summation

$$\sum_{0 \leq m \leq n \leq K_1}$$

is split into

$$\sum_{1 \leq m \leq n \leq m+m^{1/4}} + \sum_{m+m^{1/4} \leq n \leq K_1}$$

Lemma 11 :

$$B_5 = S(B_5) + O(T^{3/8}),$$

$$\text{where } S(B_5) = 4\pi \operatorname{Re} \sum_{m=1}^K \sum_{0 \leq r \leq m^{1/4}} \exp[2\pi i(2ma + 2ra + a^2)] \\ \times \int_{r+a}^{\infty} \exp[i\pi(\theta^2 - \tfrac{1}{4})] d\theta.$$

Proof: We use Lemma P1.

Lemma 12 :

$$V_{7,2} = B_6 + R(B_6) - C_6.$$

Proof: Similar to that of Lemma 10.

Lemma 13 :

$$B_6 = S(B_6) + O(T^{3/8})$$

$$\text{where } S(B_6) = 4\pi \operatorname{Re} \sum_{n=0}^{K_1} \sum_{1 \leq r \leq n^{1/4}} \int_{r-a}^{\infty} \exp[i\pi(\theta^2 - \tfrac{1}{4})] d\theta.$$

Proof: Similar to that of Lemma 10.

Lemma 14 : Let

$$S(r, a) = \sum_{r^4 \leq m \leq K} \exp(4\pi i m a)$$

$$\text{then } S(B_5) = \operatorname{Re} \sum_{0 \leq r \leq K^{1/4}} S(r, a) \int_{r+a}^{\infty} \exp[i\pi(\theta^2 - \tfrac{1}{4})] \cdot \exp[2\pi i(a^2 + 2ra)] d\theta$$

For $a \neq \frac{1}{2}$,

$$S(B_5) = O\left(\frac{\log T}{\|2a\|}\right).$$

Proof: The proof follows after changing the order of summation over r and m and noting that $S(r, a)$ is a geometric series.

Lemma 15 :

$$S(B_6) = B_{6,1} + B_{6,2} + O(T^{3/8}),$$

$$\text{where } B_{6,1} = 4\pi(K_1 + 1) \operatorname{Re} \int_{1-a}^1 \frac{\exp[i\pi(u^2 - u - \tfrac{1}{4})]}{2 \cos \pi u} du,$$

$$\text{and } B_{6,2} = 4\pi(K_1 + 1) \operatorname{Re} \sum_{r=1}^{\infty} \int_r^{\infty} \exp[i\pi(\theta^2 - \tfrac{1}{4})] d\theta.$$

Proof:

$$S(B_6) = 4\pi \operatorname{Re} \sum_{n=0}^{K_1} \sum_{1 \leq r \leq n^{1/4}} \int_{r-a}^r \exp[i\pi(\theta^2 - \tfrac{1}{4})] d\theta \\ + 4\pi \operatorname{Re} \sum_{n=0}^{K_1} \sum_{1 \leq r \leq n^{1/4}} \int_r^{\infty} \exp[i\pi(\theta^2 - \tfrac{1}{4})] d\theta.$$

The second term gives $B_{6,2}$ modulo an error $O(T^{3/8})$. The first term is

$$4\pi \operatorname{Re} \int_{1-\alpha}^1 \frac{\exp [i\pi (u^2 - u - \frac{1}{4})]}{2 \cos \pi u} [1 + \exp [\pi i (2u + 1) R]] du, \quad R = [n^{1/4}].$$

This easily gives $B_{6,1}$.

Lemma 16: For $\alpha = \frac{1}{2}$

$$\begin{aligned} S(B_5) + S(B_6) \\ = -\frac{1}{4\sqrt{2\pi}} T^{1/2} \log T + A_3 T^{1/2} + 2\sqrt{2\pi} H(\tfrac{1}{2}) T^{1/2} + O(T^{3/8}). \end{aligned}$$

Proof:

$$S(B_5) + S(B_6) = -4 \sum_{1 \leq m \leq K} \sum_{1 \leq r \leq m^{1/4}} \left\{ \frac{1}{2r+1} \right\} + \frac{A_0}{\sqrt{2\pi}} T^{1/2} + O(T^{3/8}).$$

Now we use the well-known formula

$$\sum_{r=1}^R \frac{1}{r} = \log R + \gamma + O\left(\frac{1}{R}\right)$$

and get the result.

Lemma 17:

$$V_4 = - \sum_{\substack{m=1 \\ 1 \leq r \leq m^{1/4}}}^K \frac{\sin \pi r (1 + 2\alpha)}{r} + V_{4,2} + O(T^{3/8})$$

$$\text{and} \quad V_6 = \sum_{\substack{m=1 \\ 1 \leq r \leq m^{1/4}}}^K \frac{\sin \pi r (1 + 2\alpha)}{r} + V_{6,2} + O(T^{3/8}).$$

Proof: We split the sum over n in V_4 into

$$\sum_{m < n \leq m+m^{1/4}} + \sum_{m+m^{1/4} < n \leq K}.$$

In the first sum, $n = m + r$ gives

$$\sin \left[2\pi n^2 \log \left(\frac{n}{m} \right) + 2\pi (m - n) \alpha \right] = -\sin [\pi r (1 + 2\alpha)] + O\left(\frac{r^3}{m}\right),$$

and in the denominator

$$m^{1/2} (m + r)^{1/2} \log \left(1 + \frac{r}{m} \right)$$

is replaced by r . The second sum can be treated similarly. The above results give

Lemma 18 : For $\alpha = \frac{1}{2}$

$$\begin{aligned} I_1 = & V_1 + V_2 + 2V_3 + 2V_5 - 2V_{4,2} - 2V_{6,2} - \frac{1}{4\sqrt{2\pi}} T^{1/2} \log T \\ & + R(B_5) + R(B_6) + \frac{T^{1/2}}{\sqrt{2\pi}} \left[\frac{1}{4} \log 2\pi - 4 \log 2 + \frac{1}{2} - 2\gamma \right] \\ & - C_5 - C_6 + O(T^{3/8}) \end{aligned}$$

and for $\alpha \neq \frac{1}{2}$

$$\begin{aligned} I_1 = & V_1 + V_2 + 2V_3 + 2V_5 - 2V_{4,2} - 2V_{6,2} + S(B_6) - C_5 - C_6 \\ & + R(B_5) + R(B_6) + O(\log T \parallel 2\alpha \parallel) + O(T^{3/8}). \end{aligned}$$

Proof of Theorem 3' : It is easy to see that

$$V_1 + V_2 = 2T \log K + T[C(\alpha) + \gamma - 1] - \frac{1}{\alpha} + O(1).$$

Now we only have to sum up the results of Lemmas 1 through 18 and note that for $\alpha \neq \frac{1}{2}$

$$\begin{aligned} S(B_6) = & 2\sqrt{2\pi} T^{1/2} \left[\frac{1}{2} \int_{(1-\alpha)}^1 \frac{\cos(u^2 - u - \frac{1}{4})\pi}{\cos \pi u} du \right. \\ & \left. + \sum_{r=1}^{\infty} \int_r^{\infty} \cos \pi \left(u^2 - \frac{1}{4} \right) du \right], \end{aligned}$$

and Theorem 3' is proved.

To deduce Theorem 3 from Theorem 3', we need the following lemmas.

Lemma 19 :

$$V_3 = O(T^{5/12} \log^2 T), \quad V_5 = O(T^{5/12} \log^2 T).$$

Proof : Consider V_3 and split the sum over n into

$$\sum_{m < n < 2m} + \sum_{2m \leq n \leq K}.$$

For the second sum $\log(n/m) \gg \log 2$; $\log(n/m)$ is monotonic in n for fixed m . Now consider $N < N' < 2N$, $2\pi f(x) = T \log x - 2\pi x\alpha$ and use Lemma P5. The first sum we write as

$$\sum_{K/2 < m \leq K} + \sum_{K/4 < m \leq K/2} + \dots$$

The summation is over $O(\log T)$ terms and each term is of the form

$$\sum_m \sum_{m < n < m+\lambda} + \sum_m \sum_{m+\lambda \leq n < 2m} = T_1 + T_2.$$

For T_1 , let

$$f(x) = T \log \left(\frac{x+q}{x} \right) - 2\pi q\alpha.$$

and use Lemma P4 to get

$$T_1 = O(M^{-1/2} T^{1/2} \lambda^{1/2}) + O(M^{3/2} T^{-1/2}).$$

Treating $M > T^{1/3}$ and $M \leq T^{1/3}$ separately, we get $T_1 = O(T^{5/12})$.

For T_2 ,

$$f(x) = \frac{T}{2\pi} \log x - xa,$$

and use Lemma P5 to get

$$\sum_n \frac{n^{it} \exp(-2\pi i n a)}{n^{1/2} \log\left(\frac{n}{m}\right)} = \sum_{m+\lambda < n \leq m+2\lambda} + \sum_{m+2\lambda < n \leq m+3\lambda} + \dots = O(T^{1/6} \log T).$$

Hence $T_2 = O(T^{5/12} \log T)$.

Lemma 20 :

$$C_5 + C_6 = O(T^{5/12} \log^2 T) + O\left(\frac{\log T}{a}\right).$$

Proof : Use integration by parts to get

$$\begin{aligned} C_5 + C_6 &= 2 \sum_{m,n} \frac{\sin \left[T \log \left(\frac{T}{2\pi mn'} \right) - T - \frac{\pi}{4} + 2\pi ma \right]}{m^{1/2} n'^{1/2} \log \left(\frac{T}{2\pi mn'} \right)} \\ &\quad + 2 \sum_{m,n} \frac{1}{m^{1/2} n'^{1/2}} \int_T^\infty \frac{\sin [\phi(t) - t \log mn'] \phi''(t)}{\left\{ \log \left(\frac{t}{2\pi mn'} \right) \right\}^2} dt. \end{aligned}$$

where $n' = n + a$.

Note that $\phi''(t) = t^{-1}$ and the second term on the right side is $O(\log^2 T)$. For the first term we proceed as in Lemma 19 and note that the first term is

$$\sum_{m < n < 2m} + \sum_{2m < n} + \sum_{n < m < 2n} + \sum_{2n < m} + \sum_{n=m}.$$

Each of the first four summations, treated separately, give a term $O(T^{5/12} \log^2 T)$ and the last one is $O(\log T/a)$.

Remark : It is not possible to apply the multiple average integral method of Balasubramanian [1] to $C_5 + C_6$ to reduce the error term to $O(T^{1/3})$, though this method applied to V_3 and V_4 does work and gives error terms $O(T^{1/3})$.

Lemma 21 :

$$R(B_5) + R(B_6) + C_1 + C_2 + C_3 + C_4 - 2V_{4,2} - 2V_{6,2} = O(T^{3/8}).$$

Proof : Using simple trigonometric formulae and using integration by parts, we get

$$\begin{aligned}
 C_1 &= 2 \sum_{n+n^{1/2} < m} \frac{\cos \frac{\pi}{8}(4a^2 - 5)}{\cos \pi a} \\
 &\quad \times \frac{\sin \left[2\pi(m+a)^2 \log \left(\frac{n+a}{m+a} \right) - \pi \frac{a^2}{2} + \pi a + \frac{5}{8} \pi \right]}{(m+a)^{1/2} (n+a)^{1/2} \log \left(\frac{n+a}{m+a} \right)} \\
 &\quad - 2 \sum_{n+n^{1/2} < m} \frac{\cos \frac{\pi}{8}(4a^2 - 5)}{\cos \pi a} \\
 &\quad \times \frac{\sin \left[2\pi m^2 \log \left(\frac{n+a}{m} \right) + \pi \frac{a^2}{2} - \pi a + \frac{5}{8} \pi - 2\pi m a \right]}{m^{1/2} (n+a)^{1/2} \log \left(\frac{n+a}{m+a} \right)} \\
 &\quad + O(T^{3/8}) = C_{1,1} + C_{1,2} + O(T^{3/8}).
 \end{aligned}$$

On the same lines we get similar expressions for C_2 , C_3 and C_4 . Then we can see that

$$C_{1,1} + C_{2,2} = 2V_{6,2} \quad \text{and} \quad C_{3,2} + C_{4,1} = 2V_{4,2}.$$

Also, integration by parts shows that

$$C_{1,2} + C_{2,1} + R(B_6) = O(\log^2 T),$$

$$\text{and} \quad C_{3,1} + C_{4,2} + R(B_5) = O(\log^2 T).$$

Proof of Theorem 3 : In view of Lemma 21, it only remains to show that $A_1(a) < -\sqrt{2\pi} H(a)$.

$$\begin{aligned}
 A_1(a) &= -2\sqrt{2\pi} H(a) + \sqrt{2\pi} \int_{(1-a)}^1 \frac{\cos(u^2 - u - \frac{1}{4})\pi}{\cos \pi u} du \\
 &\quad + 2\sqrt{2\pi} \sum_{r=1}^{\infty} \int_r^{\infty} \cos \pi \left(u^2 - \frac{1}{4} \right) du.
 \end{aligned}$$

For $0 \leq a \leq \frac{1}{2}$,

$$\int_{(\frac{1}{2}-a)}^{(\frac{1}{2}+a)} \frac{\cos(u^2 - u - \frac{1}{4})\pi}{\cos \pi u} du = 0.$$

Also, it is not difficult to prove that for $\alpha \geq \frac{1}{2}$

$$\int_{(1-\alpha)}^1 \frac{\cos(u^2 - u - \frac{1}{4})\pi}{\cos \pi u} du < 0 \quad (1)$$

and for $0 < \alpha \leq 1$

$$-H(\alpha) + 2 \sum_{r=1}^{\infty} \int_r^{\infty} \cos \pi \left(u^2 - \frac{1}{4} \right) du < 0. \quad (2)$$

This completes the proof of Theorem 3.

Acknowledgements

The author is indebted to Professor K Ramachandra and Mr Balasubramanian for suggesting the problem and for constant encouragement. Special thanks are due to Mr Balasubramanian for checking the manuscript and making some useful suggestions.

References

- [1] Balasubramanian R 1978 *Proc. London Math. Soc.* p. 540
- [2] Good A 1977 *Comment. Math. Helvetici* p. 52
- [3] Heath-Brown D R 1979 *Proc. London Math. Soc.* p. 385
- [4] Ramachandra K 1979 *J. Number Theory* 11 p. 465
- [5] Rane V V 1980 *J. London Math. Soc.* 21 p. 203
- [6] Titchmarsh E C 1951 *The theory of Riemann zeta function* (Oxford : Clarendon Press)
- [7] Titchmarsh E C 1934 *Q. J. Math. Oxford* 5 195

Note added in proof

In January 1981, K Ramachandra and R Balasubramanian have considered these problems in a paper entitled 'A hybrid version of a theorem of Ingham'. Their approach is simpler but their results are weaker than those proved here.

On a generalization of the class of functions with bounded Mocanu variation

J THANGAMANI

The Ramanujan Institute, University of Madras, Madras 600 005, India

MS received 11 October 1980

Abstract. The object of this paper is to generalise the well-known class of functions analytic in the unit disc having bounded Mocanu variation. Certain properties of this more general class are investigated using convolution techniques.

Keywords. Convolution ; Mocanu variation ; bounded argument rotation
 bounded boundary rotation ; hypergeometric function.

Introduction

Let M denote the class of functions $f(z)$ analytic in the unit disc $E(|z| < 1)$ with $f(0) = f'(0) - 1 = 0$. We represent by $(f * g)(z)$ the 'convolution' or the 'Hadamard product' of two functions belonging to M . Let $K_a(z)$ denote the function $z(1-z)^{-a}$ where $a \in \mathbb{C}$ with $\operatorname{Re} a > 0$. Clearly $K_a \in M$.

In this paper we introduce and study a new class of functions $MV[a, k, a]$ defined as follows :

Let $a, \alpha \in \mathbb{R}$, $a > 0$, $\alpha \geq 0$. We say $f \in MV[a, k, a]$ if and only if $f \in M$;

$$\frac{(K_{a+1} * f)(z)}{z} \neq 0; \quad \frac{(K_a * f)(z)}{z} \neq 0 \quad \text{for } z \in E \text{ and}$$

$$\int_0^{2\pi} \left| \operatorname{Re} \left\{ \alpha(a+1) \left[\frac{(K_{a+2} * f)(re^{i\theta})}{(K_{a+1} * f)(re^{i\theta})} - \frac{a}{a+1} \right] \right. \right. \\ \left. \left. + (1-\alpha)a \left[\frac{(K_{a+1} * f)(re^{i\theta})}{(K_a * f)(re^{i\theta})} - \frac{a-1}{a} \right] \right\} \right| d\theta \leq k\pi$$

where $z = re^{i\theta}$ and $k \geq 2$. (1)

It is easy to observe that when $a = 1$, $MV[a, k, a]$ coincides with the well-known class $MV[a, k]$ consisting of functions having a bounded Mocanu variation introduced by Coonce and Ziegler [1]. Similarly we see that $MV[1, 2, 1] \equiv K$

the class of convex functions; $MV[0, 2, 1] \equiv S^*$ the class of starlike functions; $MV[a, 2, 1] \equiv S_a$ the class of a -convex functions introduced by Mocanu [4]; $MV[0, k, 1] \equiv U_k$ the class of normalized functions having bounded argument rotation (Tammi [6]) and $MV[1, k, 1] \equiv V_k$ the class of normalized functions having bounded boundary rotation (Löwner [2]).

2. Geometric interpretation and preliminary remarks

A geometric interpretation of (1) can be obtained and the class $MV[a, k, a]$ can be defined in a manner analogous to $MV[a, k]$. Let

$$J(a, f, a) \equiv a(a+1) \left[\frac{(K_{a+2} * f)(z)}{(K_{a+1} * f)(z)} - \frac{a}{a+1} \right] + (1-a)a \left[\frac{(K_{a+1} * f)(z)}{(K_a * f)(z)} - \frac{a-1}{a} \right]. \quad (2)$$

If $z = re^{i\theta}$ and χ is the angle defined by

$$\chi = (1-a) \arg(K_a * f)(z) + a \arg(iK_{a+1} * f)(z),$$

using the relation

$$z(K_a * f)'(z) = a(K_{a+1} * f)(z) - (a-1)(K_a * f)(z), \quad (3)$$

it can be shown that $\partial\chi/\partial\theta = \operatorname{Re}\{J(a, f, a)\}$. Thus $f \in MV[a, k, a]$ if and only if the angle χ defined above has a bounded variation as $\arg z = \theta$ increases continuously on $|z| = r$. When $a = 1$, χ is the Mocanu angle given by $\chi = (1-a) \arg f(z) + a \arg(izf'(z))$.

Next we proceed to give an integral representation to members in the class $MV[a, k, a]$. We require the following results.

Let M_k denote the class of real valued functions $m(t)$ of bounded variation on $[-\pi, \pi]$ which satisfy the conditions

$$\int_{-\pi}^{\pi} dm(t) = 2, \quad \int_{-\pi}^{\pi} |dm(t)| \leq k.$$

The class P_k introduced by Pinchuk [5] consists of analytic functions

$$p(z) = 1 + \sum_{n=1}^{\infty} a_n z^n \text{ in } E \text{ satisfying for every}$$

$$r < 1 (z = re^{i\theta} \in E), \quad \int_0^{2\pi} |\operatorname{Re} p(z)| d\theta \leq k\pi, k \geq 2.$$

Clearly P_2 is the class P of functions with positive real part in E . It is known [5] that functions in the class P_k have the representation,

$$f(z) = \frac{1}{2} \int_{-\pi}^{\pi} \frac{1 + ze^{-it}}{1 - ze^{-it}} dm(t), \quad m(t) \in M_k, \quad (4)$$

and functions in the class U_k can be written in the form

$$f(z) = z \exp \int_{-\pi}^{\pi} -\log(1 - ze^{-it}) dm(t), \quad m(t) \in M_k. \quad (5)$$

3. Theorems and their proofs

Theorem 1. $f \in MV[a, k, a]$ if and only if

$$\alpha z \frac{(K_{a+1} * f)'(z)}{(K_{a+1} * f)(z)} + (1 - \alpha) \frac{z(K_a * f)'(z)}{(K_a * f)(z)} \in P_k. \quad (6)$$

Proof. Using (3) it is easy to check that $J(a, f, a)$ given by (2) is also equal to the left side expression given by (6). Hence the theorem follows from (1) and definition of P_k .

Corollary. Putting $\alpha = 0$ we see $f \in MV[0, k, a]$ if and only if $(K_a * f) \in U_k$ and similarly $f \in MV[1, k, a]$ if and only if $(K_{a+1} * f) \in U_k$. It follows therefore that $f \in MV[1, k, a]$ if and only if $(K_a^{-1} * K_{a+1}) * f \in MV[0, k, a]$. For $a = 1$ we obtain the well-known result [5] $f \in V_k$ iff $zf' \in U_k$.

Theorem 2. $f(z)$ is in $MV(a, k, a)$ if and only if there is an $F(z)$ in U_k such that for a suitable branch

$$(K_a * f)(z) = \left[\frac{a}{z} z^{-(a-1)/a} \int_0^z t^{-1+(a-1)/2a} F(t)^{(a+1)/2a} dt \right]^a, \quad z \in E. \quad (7)$$

Proof. Using (6) we obtain

$$\alpha \frac{(K_{a+1} * f)'(z)}{(K_{a+1} * f)(z)} + (1 - \alpha) \frac{(K_a * f)'(z)}{(K_a * f)(z)} - \frac{1}{z} = \frac{p(z) - 1}{z}$$

where $p(z) \in P_k$.

Integration and simplification making use of (4) and (5) yield

$$(K_{a+1} * f)(z) \{(K_a * f)(z)\}^{1/a-1} z^{(a-1)/2a} = F(z)^{(a+1)/2a} \quad (8)$$

where $F(z) \in U_k$. This on further simplification becomes

$$\frac{d}{dz} \left\{ (K_a * f)(z)^{1/a} z^{(a-1)/a} \right\} = \frac{a}{z} z^{(a-1)/2a-1} F(z)^{(a+1)/2a}, \quad (9)$$

from which we derive (7). The converse can be verified easily.

Corollary. Since $G(z) \in V_k$ if and only if $zG'(z) \in U_k$ we can write (7) also in the form

$$(K_a * f)(z) = \left[\frac{a}{z} z^{-(a-1)/a} \int_0^z G'(t)^{(a+1)/2a} t^{a/a-1} dt \right]^a$$

where $G(t) \in V_k$.

(10)

Remark. Putting $a = 1$ we obtain the representation for functions with bounded Mocanu variation (Goonce and Ziegler [1]). When $a = 1$, $k = 2$ we obtain the representation for α -convex functions (Mocanu [4]). When $a = 1$, $a = 1$, the functions satisfying (7) are all known to be in V_k .

The representation formula (7) enables us to obtain a distortion theorem concerning functions in $MV[a, k, a]$. We require the following lemma.

Lemma. If $F(z)$ is in U_k and $|z| = r$ then

$$r(1-r)^{(k-2)/2}(1+r)^{-(k+2)/2} \leq |F(z)| \leq r(1+r)^{(k-2)/2}(1-r)^{-(k+2)/2}. \quad (11)$$

The proof of the lemma can be found in [1].

Theorem 3. If $f(z)$ is in $MV[a, k, a]$ and $M(r) = \max |(K_a * f)re^{i\theta}|$ ($0 \leq \theta \leq 2\pi$), then

$$M(r) = 0 [(1-r)^{-(k+2)(a+1)-4a/4}] \text{ for } 0 \leq a < \frac{(k+2)(a+1)}{4}, \quad (12)$$

$$M(r) = 0 [\log(1-r)^{-1}]^a \text{ for } a = \frac{(k+2)(a+1)}{4} \quad (13)$$

$$M(r) \leq \left[\frac{a}{\alpha} 2^{(k-2)(a+1)/4\alpha} r^{-(a-1)/\alpha} \frac{\Gamma(a/\alpha) \Gamma(1 - (k+2)(a+1)/4\alpha)}{\Gamma(a/\alpha + 1 - (k-2)(a+1)/4\alpha)} \right]^a, \\ \text{for } a > \frac{(k+2)(a+1)}{4}. \quad (14)$$

Proof. By Theorem 2, there exists $F(z)$ in U_k such that

$$(K_a * f)(z) = \left[\frac{a}{\alpha} z^{-(a-1)/\alpha} \int_0^r t^{-1+(a-1)/2\alpha} F(t)^{a+1/2\alpha} dt \right]^a,$$

and applying (11) to the above we obtain

$$|(K_a * f)(re^{i\theta})|^{1/\alpha} \leq \frac{a}{\alpha} r^{-(a-1)/\alpha} \int_0^r \rho^{a/\alpha-1} (1-\rho)^{-(k+2)(a+1)/4\alpha} \\ \times (1+\rho)^{(k-2)(a+1)/4\alpha} d\rho \\ \leq \frac{a}{\alpha} r^{-(a-1)/\alpha} 2^{(k-2)(a+1)/4\alpha} \int_0^r \rho^{a/\alpha-1} (1-\rho)^{-(k+2)(a+1)/4\alpha} d\rho.$$

Changing the variable by putting $\rho = ru$ we obtain

$$|(K_a * f)(re^{i\theta})|^{1/\alpha} \leq \frac{a}{\alpha} r^{1/\alpha} 2^{(k-2)(a+1)/4\alpha} \int_0^1 u^{a/\alpha-1} (1-ru)^{-(k+2)(a+1)/4\alpha} du. \quad (15)$$

using hypergeometric functions given by

$$\begin{aligned} G(l, m, n; z) &= \frac{\Gamma(n)}{\Gamma(l)\Gamma(m)} \sum_{p=0}^{\infty} \frac{\Gamma(l+p) \Gamma(m+p)}{\Gamma(n+p)} \frac{z^p}{p!} \\ &= \frac{\Gamma(n)}{\Gamma(l)\Gamma(n-l)} \int_0^1 u^{l-1} (1-u)^{n-l-1} (1-zu)^{-m} du, \end{aligned} \quad (16)$$

where $\operatorname{Re} l > 0$ and $\operatorname{Re} (n-l) > 0$ we can write

$$|(K_a * f)(re^{i\theta})|^{1/a} \leq r^{1/a} 2^{(k-2)(a+1)/4a} G\left(\frac{a}{a}, \frac{(k+2)(a+1)}{4a}, 1 + \frac{a}{a}; r\right). \quad (17)$$

$$\text{Hence } M(r) \leq r 2^{\frac{1}{4}[(k+2)(a+1)]} \left[G\left(\frac{a}{a}, \frac{1}{4a} [(k+2)(a+1)], 1 + \frac{a}{a}; r\right) \right]^a \quad (18)$$

If we now take $0 < a < \frac{1}{4}[(k+2)(a+1)]$ then

$$\begin{aligned} \lim_{r \rightarrow 1-0} \frac{G\left(\frac{a}{a}, \frac{1}{4a} [(k+2)(a+1)], 1 + \frac{a}{a}; r\right)}{(1-r)^{1-\frac{1}{4a}[(k+2)(a+1)]}} \\ = \frac{4a}{(k+2)(a+1) - 4a} \quad ([7] \text{ p. 299}). \end{aligned}$$

Combining this with (18) we get

$$M(r) = O\{r(1-r)^{-\frac{1}{4}[(k+2)(a+1)-4a]}\} = O\{(1-r)^{-\frac{1}{4}[(k+2)(a+1)-4a]}\} \text{ as } r \rightarrow 1-0$$

when $a = \frac{1}{4}[(k+2)(a+1)]$ we obtain from ([7], p. 299)

$$\lim_{r \rightarrow 1-0} \frac{G\left(\frac{a}{a}, 1, 1 + \frac{a}{a}; r\right)}{\log[(1-r)^{-1}]} = a/a,$$

and combining with (18) we have

$$M(r) = O\{r[\log(1-r)^{-1}]^a\} = O[\log(1-r)^{-1}]^a \text{ as } r \rightarrow 1-0.$$

In the case $a > [(k+2)(a+1)]/4$ we have

$$\int_0^1 \rho^{\frac{a}{a}-1} (1-\rho)^{-\frac{1}{4a}[(k+2)(a+1)]} d\rho = \frac{\Gamma(a/a) \Gamma\left(1 - \frac{1}{4a}[(k+2)(a+1)]\right)}{\Gamma\left(\frac{a}{a} + 1 - \frac{1}{4a}[(k+2)(a+1)]\right)}.$$

Hence we obtain

$$M(r) \leq \left[\frac{a}{\alpha} 2^{\frac{1}{4\alpha} [(k-2)(a+1)]} r^{-\frac{(a-1)}{\alpha}} \frac{\Gamma(a/\alpha) \Gamma\left(1 - \frac{1}{4\alpha} [(k+2)(a+1)]\right)}{\Gamma\left(\frac{a}{\alpha} + 1 - \frac{1}{4\alpha} [(k+2)(a+1)]\right)} \right]^\alpha$$

which completes the proof of the theorem.

The function $f(z)$ defined by

$$(K_a * f)(z) = \frac{a}{\alpha} z^{-(a-1)\alpha} \int_0^z t^{a/\alpha-1} (1-t)^{-\frac{1}{4\alpha} [(k+2)(a+1)]} (1+t)^{\frac{1}{4\alpha} [(k-2)(a+1)]} dt$$

shows that the orders given in (12) and (13) are the best possible.

Remark. When $a = 1$ we obtain the results in [1] and when $a = 1$, $k = 2$ the results in [3].

It is worth noting that when $a = 1$, $MV[a, k, a] \subset S$, the class of univalent functions iff $k \leq 2 + 2a$ and $MV[a, k, a] \subset MV[\beta, k, a]$ if $a \geq \beta \geq 0$; [1]. These results are under investigation by the author for other values of $a \in \mathbf{R}$.

Acknowledgement

The author is thankful to Prof. K S Padmanabhan for his help and suggestions in the preparation of this paper.

References

- [1] Coonce H B and Ziegler M R 1974 *Rev. Roum. Math. Pures Appl.* **19** 1093
- [2] Löwner K 1917 *Ber. der Kon. Sächsischen Gesellschaft der Wiss Zu Leipzig* 26
- [3] Miller S S 1973 *Proc. Am. Math. Soc.* **38** 311
- [4] Mocanu P T 1969 *Math. Cluj.* **11** 127
- [5] Pinchuk B 1971 *Isr. J. Math.* **10** 6
- [6] Tammi O 1952 *Ann. Acad. Sci. Fennicae Ser. A.I. Math. Phys.* **114** 51
- [7] Whittaker E T and Watson G N 1927 *A course of modern analysis* (Cambridge Univ. Press)

Analyse asymptotique des équations de Transport dans le cas d'évolution

RAFAEL F SANTOS

Departement de Mathematiques, Université de Paris XIII, 93430 Villetaneuse, France

MS received 15 December 1980

Abstract. Nous démontrons la convergence de la solution d'une équation de Transport vers la solution d'une équation de Diffusion quand le libre parcours moyen tend vers zéro, pour des équations d'évolution et un domaine borné dans deux cas: flux incident nul et réflexion spéculaire.

Keywords. Transport equations; diffusion approximation; Fredholm alternative; boundary layers.

1. Introduction

Soit X un ouvert régulier dans R^3 et S^2 la sphère unitaire. On désigne par ∂X la frontière de X et par $\nu = \nu(x)$ la normale extérieure à X , au point $x \in \partial X$.

Considérons l'équation de Transport, dans $[0, T] \times X \times S^2$,

$$(1) \quad \frac{\partial u^\epsilon}{\partial t} + \frac{v \cdot \nabla u^\epsilon}{\epsilon} + \frac{(1 - \epsilon^2 \gamma)}{\epsilon^2} \sigma(x) u^\epsilon - \frac{\sigma(x)}{\epsilon^2} \int_{S^2} f(x, v, v') u^\epsilon(t, x, v') dv' = 0$$

et soit $\Gamma^- = \{(x, v) \in \partial X \times S^2 : v \cdot \nu < 0\}$.

On se propose, dans les conditions que l'on précisera dans la suite, de construire un développement asymptotique de u^ϵ , dans un voisinage de $\epsilon = 0$.

Ceci a été fait dans l'espace entier par Benssoussan, Lions et Papanicolaou [1]. Dans le cas d'un ouvert borné, donc d'un problème aux limites, ces auteurs n'ont étudié que le problème stationnaire, γ étant de signe constant lorsqu'ils considèrent sur le bord des conditions de réflexion. Nous étudions le cas d'évolution avec deux types de conditions aux limites : absence de flux rentrant et réflexion spéculaire. Alors nous pouvons prendre γ positif ou négatif.

Dans le paragraphe 2 on considère que la condition aux limites est $u^\epsilon|_{\Gamma^-} = 0$ (c'est-à-dire, une absence de flux incident). Pour ce problème on construit le terme d'ordre zéro comme solution d'une équation de Diffusion pulle sur le bord.

Pour le correcteur d'ordre un, outre un terme à l'intérieur on a besoin de deux termes de plus : un de couche initiale, que l'on obtient en étendant le temps et un autre de couche limite que l'on obtient en étendant les coordonnées spatiales au voisinage du bord. Un bon choix de la donnée initiale nous permet de se passer d'un éventuel terme de couche à la fois initiale et limite.

Au paragraphe 3 on montre que lorsque l'on considère la réflexion spéculaire, le premier terme du développement vérifie une équation de Diffusion avec une condition de Neumann sur le bord.

Dans toute la suite nous supposons que

- (i) σ est une fonction suffisamment régulière dans X et vérifie $0 < m \leq \sigma \leq M$.
- (ii) f est une fonction mesurable, bornée, suffisamment régulière en $x \in X$, strictement positive et vérifie la condition de normalisation

$$\int_{S^2} f(x, v, v') dv' = \int_{S^2} f(x, v, v) dv = 1.$$

On pose K l'opérateur défini par

$$K\varphi = \int_{S^2} f(x, v, v') \varphi(v') dv' \text{ et}$$

$$Q\varphi = \sigma [I - K] \varphi.$$

On note $\|\cdot\|_0$ la norme dans $L^\infty(X \times S^2)$ et pour $g, h \in L^2(S^2)$ on pose

$$(g, h) = \frac{1}{4\pi} \int_{S^2} g(v) h(v) dv.$$

2. On suppose que σ et f ne dépendent que de la première composante de x , que l'on note encore x , et que $x \in [0, 1]$ (à fin de simplifier les notations de couche limite on prend pour X une bande, cf. [4]). Donc

$$\Gamma^- = \{(0, v) : v_1 > 0\} \cup \{(1, v) : v_1 < 0\} \equiv \Gamma_0^- \cup \Gamma_1^-.$$

On note T_ϵ l'opérateur défini par

$$D(T_\epsilon) = \left\{ u \in C(X \times S^2) : v_1 \frac{\partial u}{\partial x} \in C(X \times S^2) \right\},$$

$$T_\epsilon u = \frac{v_1}{\epsilon} \frac{\partial u}{\partial x} + \frac{Qu}{\epsilon^2} - \gamma \sigma u.$$

D'après l'interprétation probabiliste des processus de Transport et suivant [2] on a le

LEMME 1 : La solution de

$$\frac{du}{dt} + T_\epsilon u = p$$

$$u|_{\Gamma^-} = g$$

$$u(0) = h$$

vérifie $\|u(t)\|_0 \leq e^{\sup(\gamma\sigma)t} \sup(\|p\|_0, \|g\|_0, \|h\|_0).$

THEOREME 1 : Soient φ_0, φ_1 telles que $\varphi_0 \in C_c^\infty(X)$, $\varphi_1(\cdot, v) \in C^\infty(\bar{X})$, $\varphi_1(\cdot, v)|_{\partial x} = 0$, $\varphi_1(x, \cdot) \in L^1(S^2)$. Alors la solution de

$$(2) \quad \frac{du^\epsilon}{dt} + T_\epsilon u^\epsilon = 0,$$

$$(3) \quad u^\epsilon|_{\Gamma^-} = 0,$$

$$(4) \quad u^\epsilon(0) = \varphi_0(x) + \epsilon \varphi_1(x, v),$$

vérifie
$$u^\epsilon(t, x, v) = u_0(t, x) + \epsilon \left[u_1(t, x, v) + u_1^I\left(\frac{t}{\epsilon^2}, x, v\right) + \sum_{\alpha=0}^1 u_{1,\alpha}^{CL}(t, \xi_\alpha, v) \right] + \omega_\epsilon, \quad \xi_0 = \frac{x}{\epsilon}, \quad \xi_1 = \frac{1-x}{\epsilon}.$$

De plus, il existe une constante $C > 0$, indépendante de ϵ , telle que, pour $t \in [0, T]$, on ait $\|\omega_\epsilon(t)\|_0 \leq \epsilon^2 C e^{\gamma_0 t}$, $\gamma_0 = \sup(\sup(\gamma\sigma), 0)$.

Démonstration : On écrit

$$u^\epsilon(t, x, v) = \psi^\epsilon\left(t, \frac{t}{\epsilon^2}, x, \xi_0, \xi_1, v\right) + u^I\left(\frac{t}{\epsilon^2}, x, v\right) + u_0^{CL}(t, \xi_0, v) + u_1^{CL}(t, \xi_1, v)$$

avec $\psi^\epsilon \in C(\mathbf{R}_+^2 \times [0, 1] \times \mathbf{R}^2 \times S^2)$, $u^I \in C(\mathbf{R}_+ \times [0, 1] \times S^2)$,

$$u_\alpha^{CL} \in C(\mathbf{R}_+ \times \mathbf{R}_+ \times S^2)$$

et on pose

$$(5) \quad \psi^\epsilon = \sum_{k=0}^3 \epsilon^k u_k + \omega_\epsilon, \quad u^I = \sum_{k=1}^3 \epsilon^k u_k^I, \quad u_\alpha^{CL} = \sum_{k=1}^3 \epsilon^k u_{k,\alpha}^{CL}.$$

Soit $\zeta = t/\epsilon^2$. Le problème (2), (3), (4) devient

$$(6.1) \quad \frac{\partial \psi^\epsilon}{\partial t} + \frac{1}{\epsilon^2} \frac{\partial \psi^\epsilon}{\partial \zeta} + \frac{v_1}{\epsilon} \frac{\partial \psi^\epsilon}{\partial x} + \frac{v_1}{\epsilon^2} \left[\frac{\partial \psi^\epsilon}{\partial \xi_0} - \frac{\partial \psi^\epsilon}{\partial \xi_1} \right] + \frac{1}{\epsilon^2} Q \psi^\epsilon - \gamma \sigma \psi^\epsilon = 0,$$

$$(6.2) \quad \frac{1}{\epsilon^2} \frac{\partial u^I}{\partial \zeta} + \frac{v_1}{\epsilon} \frac{\partial u^I}{\partial x} + \frac{1}{\epsilon^2} Q u^I - \gamma \sigma u^I = 0,$$

$$(6.3) \quad \frac{\partial u_\alpha^{CL}}{\partial t} + (-1)^\alpha \frac{v_1}{\epsilon^2} \frac{\partial u_\alpha^{CL}}{\partial \xi_\alpha} + \frac{1}{\epsilon^2} Q^\alpha u_\alpha^{CL} - \gamma \sigma u_\alpha^{CL} = 0,$$

$$(7.1) \quad \psi^\epsilon(t, \zeta, 0, 0, 1/\epsilon, v) + u^I(\zeta, 0, v) + u_0^{CL}(t, 0, v) + u_1^{CL}(t, 1/\epsilon, v) = 0,$$

$$(7.2) \quad \psi^\epsilon(t, \zeta, 1, 1/\epsilon, 0, v) + u^I(\zeta, 1, v) + u_0^{CL}(t, 1/\epsilon, v) + u_1^{CL}(t, 0, v) = 0,$$

$$(8) \quad \psi^\epsilon(0, 0, x, \xi_0, \xi_1, v) + u^I(0, x, v) + u_0^{CL}(0, \xi_0, v) + u_1^{CL}(0, \xi_1, v) = \varphi_0(x) + \epsilon \varphi_1(x, v),$$

où Q^α est l'opérateur Q lorsque $x = \xi_\alpha$. On reporte (5) dans (6). A l'intérieur on obtient les équations

$$(9) \quad Q u_0 = 0,$$

$$(10) \quad Qu_1 = -v \frac{\partial u_0}{\partial x},$$

$$(11) \quad Qu_2 = - \left[\frac{\partial u_0}{\partial t} + v_1 \frac{\partial u_1}{\partial x} - \gamma \sigma u_0 \right],$$

$$(12) \quad Qu_3 = - \left[\frac{\partial u_1}{\partial t} + v_1 \frac{\partial u_2}{\partial x} - \gamma \sigma u_1 \right].$$

Compte tenu des hypothèses sur f (positivité et condition de normalisation) on déduit de l'équation (9) que u_0 ne dépend pas de v : $u_0 = u_0(t, x)$. D'autre part, 1 est la mesure sur S^2 invariante pour le semigroupe engendré par Q . Ainsi, en vertu de l'alternative de Fredholm, on déduit que

$$u_1 = D(x, v) \frac{\partial u_0}{\partial x} + c_1(t, x)$$

où $D(x, v)$ est défini par $QD + v_1 = 0$, $(D, 1) = 0$. En appliquant à nouveau l'alternative de Fredholm on montre qu'il faut et qu'il suffit que $u_0(t, x)$ et $c_1(t, x)$ vérifient dans $[0, T] \times X$, respectivement,

$$(13) \quad \frac{\partial u_0}{\partial t} + \frac{\partial}{\partial x} \left[(v_1, D(x, v)) \frac{\partial u_0}{\partial x} \right] - \gamma \sigma u_0 = 0$$

$$(14) \quad \frac{\partial c_1}{\partial t} + \frac{\partial}{\partial x} \left[(v_1, D(x, v)) \frac{\partial c_1}{\partial x} \right] - \gamma \sigma c_1 = - \left(v_1, \frac{\partial g}{\partial x} \right)$$

(où $g(t, x, v)$ vérifie

$$Qg = - \left[\frac{\partial u_0}{\partial t} + v_1 \frac{\partial}{\partial x} D(x, v) \frac{\partial u_0}{\partial x} - \gamma \sigma u_0 \right]$$

pour que les équations (11) et (12) aient une solution et

$$u_2 = g(t, x, v) + D(x, v) \frac{\partial c_1}{\partial x} + c_2(t, x).$$

Pour avoir les conditions aux limites et initiale pour $u_0(t, x)$ on tient compte des équations (7) et (8). On en déduit que l'on doit avoir

$$(15) \quad u_0(t, x) = 0, \text{ sur } [0, T] \times \partial X,$$

$$(16) \quad u_0(t, x) = \varphi_0(x), \text{ dans } X,$$

Comme l'opérateur

$$\frac{\partial}{\partial x} \left[(v_1, D(x, v)) \frac{\partial}{\partial x} \right]$$

est uniformément elliptique [3], il existe une unique u_0 , suffisamment régulière, solution de (13), (15), (16). Les conditions aux limites et initiale pour $c_1(t, x)$ seront obtenues d'après les termes $u_{1,\alpha}^{CL}$ et u_1^I .

Pour les termes de couche initiale on obtient de (6.2) les équations

$$(17) \quad \frac{\partial u_1^I}{\partial \zeta} + Qu_1^I = 0,$$

$$(18) \quad \frac{\partial u_2^I}{\partial \zeta} + Qu_2^I = -v_1 \frac{\partial u_1^I}{\partial x},$$

$$(19) \quad \frac{\partial u_3^I}{\partial \zeta} + Qu_3^I = -v_1 \frac{\partial u_2^I}{\partial x} + \gamma \sigma u_1^I.$$

Outre (17) u_1^I doit vérifier, à l'instant zéro

$$(20) \quad u_1^I(0, x, v) + u_1(0, x, v) = \phi_1(x, v),$$

et la condition de couche initiale

$$(21) \quad u_1^I(\zeta, \cdot, \cdot) \rightarrow 0, \quad \zeta \rightarrow +\infty.$$

D'autre part, pour les termes de couche limite on obtient

$$(22) \quad (-1)^a v_1 \frac{\partial u_{1,\alpha}^{CL}}{\partial \xi_a} + Q^a u_{1,\alpha}^{CL} = 0,$$

$$(23) \quad (-1)^a v_1 \frac{\partial u_{2,\alpha}^{CL}}{\partial \xi_a} + Q^a u_{2,\alpha}^{CL} = 0,$$

$$(24) \quad (-1)^a v_1 \frac{\partial u_{3,\alpha}^{CL}}{\partial \xi_a} + Q^a u_{3,\alpha}^{CL} = \gamma \sigma u_{1,\alpha}^{CL} - \frac{\partial u_{1,\alpha}^{CL}}{\partial t},$$

les $u_{1,\alpha}^{CL}$ vérifiant

$$(25) \quad u_{1,\alpha}^{CL}(t, 0, v) + u_1(t, \alpha, v) = 0,$$

$$(26) \quad u_{1,\alpha}^{CL}(\cdot, \xi_a, \cdot) \rightarrow 0, \quad \xi_a \rightarrow +\infty.$$

Les fonctions $u_{1,\alpha}^{CL}$ et u_1^I sont donc complètement déterminées si on applique les lemmes suivants ([1]) :

LEMME 2 : Soit $g(\zeta, \cdot)$ telle que $|g(\zeta, \cdot)| \leq C_1 \exp(-\delta\zeta)$. Alors pour toute constante C il existe u , solution unique de

$$\frac{du}{d\zeta} + Qu + g = 0,$$

$$u(0) = g_0 + C.$$

De plus on a $u(\zeta) \rightarrow (g_0, 1)$ si $g = 0$ et $C = 0$ et il existe une unique C telle que $u(\zeta) \rightarrow 0$, de façon exponentielle, lorsque $\zeta \rightarrow +\infty$.

LEMME 3 : (a) Il existe p_a , mesure sur Γ_a^- telle que la solution de

$$v_a \cdot v \frac{\partial u}{\partial \xi_a} + Q^a u = 0,$$

$$u(0, \cdot) |_{\Gamma_a^-} = h_0,$$

vérifie $u(\xi_a, \cdot) \rightarrow (p_a, h_0)$, de façon exponentielle, lorsque $\xi_a \rightarrow +\infty$.

(b) Si $h(\xi_a, \cdot)$ vérifie $|h(\xi_a, \cdot)| \leq C_1 \exp(-\beta\xi_a)$, alors pour toute constante C il existe u solution unique de

$$v_a \cdot v \frac{\partial u}{\partial \xi_a} + Q^a u + h = 0,$$

$$u(0, \cdot) |_{\Gamma_a^-} = h_0 + C,$$

et il existe une unique C telle que $u(\xi_a, \cdot) \rightarrow 0$, de façon exponentielle, quand $\xi_a \rightarrow +\infty$.

Ainsi, d'après le lemme 2, pour que l'on ait (21), il faut que $(\varphi_1(x, \cdot) - u_1(0, x, \cdot), 1) = 0$ c'est-à-dire

$$\left(\varphi_1(x, \cdot) - D(x, \cdot) \frac{\partial \varphi_0}{\partial x} - c_1(0, x), 1 \right) = 0.$$

Ceci détermine la condition initiale pour $c_1(t, x)$:

$$(27) \quad c_1(0, x) = (\varphi_1(x, \cdot), 1).$$

Le Lemme 3 et (26) entraînent

$$(u_1(t, a, \cdot), p_a) = \left(c_1(t, a) + D(a, \cdot) \frac{\partial u_0}{\partial x}(t, a), p_a \right) = 0,$$

donc

$$(28) \quad c_1(t, a) = - (D(a, \cdot), p_a) \frac{\partial u_0}{\partial x}(t, a),$$

et $c_1(t, x)$ est déterminé par (14), (27) et (28). Quitte à préciser les conditions initiales et aux limites les lemmes 2 et 3 nous assurent l'existence des termes d'ordre 2 et 3.

Soit $\omega^\epsilon(t, x, v) = \omega_\epsilon(t, \zeta, x, \xi_0, \xi_1, v)$; ω^ϵ est solution de

$$\begin{aligned} \frac{\partial \omega^\epsilon}{\partial t} + T_\epsilon \omega^\epsilon &= \epsilon^2 \left[-\frac{\partial u_2}{\partial t} - v_1 \frac{\partial u_3}{\partial x} + \gamma \sigma u_2 \right] + \epsilon^3 \left[-\frac{\partial u_3}{\partial t} + \gamma \sigma u_3 \right] \\ &= O(\epsilon^2) \end{aligned}$$

$$\omega^\epsilon|_{\Gamma^-} = R,$$

$$\omega^\epsilon(0) = R_0$$

$$\text{où} \quad R(a) = u^\epsilon(t, a, v) - u_0(t, a) - \epsilon [u_1(t, a, v) + u_1^I(\zeta, a, v)$$

$$+ \sum_{\alpha=0}^1 u_{1,\alpha}^{CL}(t, 0, v)] + O(\epsilon^2),$$

$$= -\epsilon u_1^I(\zeta, a, v) + O(\epsilon^2) = O(\epsilon^2).$$

$$\text{et} \quad R_0 = u^\epsilon(0, x, v) - u_0(0, x) - \epsilon [u_1(0, x, v) + u_1^I(0, x, v)$$

$$+ \sum_{\alpha=0}^1 u_{1,\alpha}^{CL}(0, \xi_a, v)] + O(\epsilon^2)$$

$$= -\epsilon \sum_{\alpha=0}^1 u_{1,\alpha}^{CL}(0, \xi_a, v) + O(\epsilon^2) = O(\epsilon^2).$$

En effet on a $u_1^I(\zeta, a, v) = 0$ et $u_{1,\alpha}^{CL}(0, \xi_a, v) = 0$.

D'après le lemme 1 on a, pour

$$t \in [0, T], \quad |\omega^\epsilon(t)|_0 \leq \epsilon^2 C \exp[\sup(\gamma\sigma)t]$$

et on conclut.

3. On suppose que X est borné et qu'il représente un milieu isotrope, donc que f est constante. On définit T_ϵ par

$$D(T_\epsilon) = \{u \in C(X \times S^2) : v \cdot \nabla u \in C(X \times S^2)\},$$

$$T_\epsilon u = \frac{v \cdot \nabla u}{\epsilon} + \frac{Qu}{\epsilon^2} - \gamma \sigma u$$

et $R : \Gamma_x^+ \rightarrow \Gamma_x^-$ par $R\phi(x, v) = \phi(x, v - (2v \cdot v)v)$.

THEOREME 2: Soit ϕ_0 une fonction régulière dans $\bar{X}$ vérifiant $\partial\phi_0/\partial\nu = 0$ sur ∂X . Alors la solution de

$$(29) \quad \frac{du^\epsilon}{dt} + T_\epsilon u^\epsilon = 0,$$

$$(30) \quad u^\epsilon(t, x, v)|_{\Gamma^-} = R[u^\epsilon(t, x, v)|_{\Gamma^+}],$$

$$(31) \quad u^\epsilon(0) = \phi_0(x),$$

vérifie $u^\epsilon(t, x, v) = u_0(t, x) + \omega^\epsilon$. De plus, pour tout $\beta > 0$, il existe une constante positive $C = C(\beta)$ indépendante de ϵ telle que pour $t \in [0, T]$, on ait

$$|\omega^\epsilon(t)|_0 \leq \epsilon C \exp[(\gamma_0 + \beta)t], \quad \gamma_0 = \sup(\sup(\gamma\sigma), 0).$$

Démonstration : Posons

$$(32) \quad u^\epsilon(t, x, v) = u_0(t, x) + \epsilon u_1(t, x, v) + \epsilon^2 u_2(t, x, v) + \omega^\epsilon,$$

et reportons (32) dans (29). On trouve

$$(33) \quad Qu = 0,$$

$$(34) \quad Qu = -v \cdot \nabla u_0,$$

$$(35) \quad Qu = -\left[\frac{\partial u_0}{\partial t} + v \cdot \nabla u_1 - \gamma \sigma u_0\right].$$

En raisonnant comme dans la démonstration du théorème 1 on déduit que u_0 ne dépend pas de v , que, d'après l'alternative de Fredholm $u_1 = -v \cdot \nabla u_0 + c_1(t, x)$ et que u_0 satisfait dans $[0, T] \times X$ l'équation

$$(36) \quad \frac{\partial u_0}{\partial t} - \frac{1}{3} \Delta u_0 - \gamma \sigma u_0 = 0,$$

u_2 étant donnée par

$$u_2 = -\frac{\partial u_0}{\partial t} + v \cdot \nabla u_1 - \gamma \sigma u_0 + c_2(t, x).$$

En reportant (32) dans (30) on trouve à l'ordre un que

$$v \cdot \nabla u_0 = (v - (2v \cdot v)v) \cdot \nabla u_0 \text{ et donc}$$

$$(37) \quad \partial u_0 / \partial \nu = 0, \text{ sur } [0, T] \times \partial X.$$

Finalement (31) entraîne

$$(38) \quad u_0(0, x) = \phi_0(x)$$

Les équations (36)–(38) déterminent u_0 de façon unique.

On a $|\omega^\epsilon|_0 \leq \omega_\epsilon$ où ω_ϵ satisfait

$$\frac{d\omega_\epsilon}{dt} + T_\epsilon \omega_\epsilon = \epsilon \left| -\frac{\partial u_1}{\partial t} - v \cdot \nabla u_2 + \gamma \sigma u_1 \right| + \epsilon^2 \left| -\frac{\partial u_2}{\partial t} + \gamma \sigma u_2 \right| \equiv h^\epsilon$$

$$h^\epsilon = O(\epsilon),$$

$$\omega_\epsilon|_{\Gamma^-} = R[\omega_\epsilon|_{\Gamma^+}] + \epsilon f^\epsilon,$$

$$f^\epsilon = O(\epsilon),$$

$$\omega_\epsilon(0) = g^\epsilon,$$

$$g^\epsilon = O(\epsilon).$$

LEMME 4 : Pour tout $\alpha > 0$ il existe une constante $C(\alpha)$ telle que la solution de

$$\frac{du}{dt} + \frac{v \cdot \nabla u}{\epsilon} + \frac{Qu}{\epsilon^2} + \alpha u = 0 \quad Q1 = 0,$$

$$u|_{\Gamma^-} = R[u|_{\Gamma^+}] + \epsilon p^\epsilon$$

$$R1 = 1,$$

$$u(0) = q^\epsilon$$

verifie $|u(t)|_0 \leq \exp(-\alpha t) |q^\epsilon|_0 + C(\alpha) |p^\epsilon|_0$.

Idee de la démonstration : on tient compte de la description probabiliste pour écrire

$$u(t, x, v) = E_{x,v} \{ e^{-\alpha t} q^\epsilon + \int_0^t e^{-\alpha s} p^\epsilon dN(s) \}$$

et on applique (1) (lemme p. 144). On en déduit le

LEMME 5 : Pour tout $\alpha > 0$ il existe une constante $C(\alpha)$ telle que la solution de

$$\frac{\partial \psi}{\partial t} + \frac{v \cdot \nabla \psi}{\epsilon} + \frac{Q\psi}{\epsilon^2} = H^\epsilon \quad Q1 = 0,$$

$$\psi|_{\Gamma^-} = R[\psi|_{\Gamma^+}] + \epsilon p^\epsilon$$

$$R1 = 1,$$

$$\psi(0) = q^\epsilon,$$

verifie $|\psi(t)|_0 \leq |q^\epsilon|_0 + t \exp(\alpha t) |H^\epsilon|_0 + C(\alpha) \exp(\alpha t) |p^\epsilon|_0$.

Or $\omega_\epsilon \leq \phi^\epsilon$ où ϕ^ϵ est la solution de

$$\frac{\partial \phi^\epsilon}{\partial t} + \frac{v \cdot \nabla \phi^\epsilon}{\epsilon} + \frac{Q\phi^\epsilon}{\epsilon^2} - \gamma_0 \phi^\epsilon = h^\epsilon,$$

$$\phi^\epsilon|_{\Gamma^-} = R[\phi^\epsilon|_{\Gamma^+}] + \epsilon f^\epsilon,$$

$$\phi^\epsilon(0) = g^\epsilon.$$

En effet, $u = \omega_\epsilon - \phi^\epsilon$ est solution de

$$\frac{du}{dt} + \frac{v \cdot \nabla u}{\epsilon} + \frac{Qu}{\epsilon^2} - \gamma \sigma u = \phi(\gamma \sigma - \gamma_0) \leq 0,$$

$$u|_{\Gamma^-} = R[u|_{\Gamma^+}],$$

$$u(0) = 0$$

d'où $u \leq 0$. Posons $W = \phi^\epsilon e^{-\gamma_0 t}$; W est solution de

$$\frac{dW}{dt} + \frac{v \cdot \nabla W}{\epsilon} + \frac{QW}{\epsilon^2} = H^\epsilon, \quad H^\epsilon = h^\epsilon e^{-\gamma_0 t},$$

$$W|_{\Gamma^-} = R[W|_{\Gamma^+}] + \epsilon F^\epsilon, \quad F^\epsilon = f^\epsilon e^{-\gamma_0 t},$$

$$W(0) = g^\epsilon,$$

et donc pour tout

$$a > 0, \quad |W(t)|_0 \leq |g^\epsilon|_0 + t e^{at} |H^\epsilon|_0 + C(a) e^{at} |F^\epsilon|_0.$$

Donc

$$\begin{aligned} \omega_\epsilon \leq \varphi^\epsilon \leq \exp(\gamma_0 t) |g^\epsilon|_0 + t \exp[\alpha + \gamma_0] t |H^\epsilon|_0 + C(a) \\ \times \exp([\alpha + \xi_0] t) |F^\epsilon|_0, \end{aligned}$$

et pour tout $\alpha' > \alpha$ il existe $C(\alpha')$ telle que

$$t \exp([\alpha + \gamma_0] t) \leq C(\alpha') \exp[\alpha' + \gamma_0] t.$$

D'où le résultat.

Pour f quelconque on aurait un résultat pareil avec une condition aux limites du type $\partial u_0 / \partial \nu_a = 0$.

References

- [1] Bensoussan A, Lions J L et Papanicolaou G 1979 *J. Publ. RIMS Kyoto Univ.* **15** 53
- [2] Blankenship G et Papanicolaou 1978 *SIAM J. Appl. Math.* **34** 437
- [3] Sentis 1979 *CR Acad. Sci.* **A289** 567
- [4] Sentis R 1981 Analyse asymptotique des équations de transport (à paraître).

Application of Newton's method to a homogenization problem

S KESAVAN

TIFR Centre, Indian Institute of Science Campus, Bangalore 560 012, India

MS received 9 October 1980 ; revised 23 December 1980

Abstract. The homogenization of a family (P_ϵ) of uniformly elliptic semilinear partial differential equations of second order is studied. The main result is that any non-singular solution u of the homogenized problem (P) is the limit of non-singular solutions of (P_ϵ) . The method consists of specifying a function w_ϵ starting from which the Newton iterates converge to a solution u_ϵ of P_ϵ . These solutions u_ϵ converge to the given solution u of (P) .

Keywords. Elliptic ; homogenization ; Newton's method ; semilinear.

1. Introduction

Let $\Omega \subset \mathbf{R}^n$ be a bounded open set with boundary Γ . Consider the following family of second-order differential operators defined in Ω , depending on a parameter ϵ which tends to zero :

$$(1.1) \quad A^\epsilon = - \frac{\partial}{\partial x_i} \left(a_{ij}^\epsilon(x) \frac{\partial}{\partial x_j} \right).$$

(Here, as well as throughout this paper, the summation convention for repeated indices will be assumed.) Various homogenization problems connected with these operators have been considered in the literature. Benscussan *et al* [1] have studied the linear problem $A^\epsilon u_\epsilon = f$ with various boundary conditions. Kesavan [4] has analysed the corresponding eigenvalue problems. In this paper the following semilinear problem will be studied :

$$(1.2) \quad A^\epsilon u_\epsilon = f(u_\epsilon) \text{ in } \Omega,$$

$$(1.3) \quad u_\epsilon = 0 \text{ on } \Gamma,$$

where $f: \mathbf{R} \rightarrow \mathbf{R}$ is a given function.

Indeed the identification of the homogenized problem is not difficult and will not occupy the principal part of this study. In fact Mythily [5] has homogenized (1.2) with a more complicated boundary condition. The more important question, which will be the main preoccupation in this paper, is the following : ' given a solution u of the homogenized problem, does it approximate a solution u_ϵ of the real problem ? ' In other words, ' is u the limit of a sequence $\{u_\epsilon\}$ of solutions ? '.

In the case of the linear problem, the uniqueness of the solutions of the various problems concerned trivially answers the above question in the affirmative. In the case of the eigenvalue problem Kesavan [4] has shown that eigenfunctions u of a *simple* eigenvalue of the homogenized problem occur as limits of eigenfunctions u_ϵ of the given family of problems. The case of multiple eigenvalues is still open. In this paper, it will be proved that, under suitable conditions, a *non-singular* solution u of the homogenized problem is the limit of a family $\{u_\epsilon\}$ of solutions of (1.2)–(1.3), which, for sufficiently small ϵ , will also be non-singular.

The main idea of the proof is the following: First a function w_ϵ is defined as the unique solution of the auxiliary linear problem:

$$(1.4) \quad A^\epsilon w_\epsilon = f(u) \text{ in } \Omega,$$

$$(1.5) \quad w_\epsilon = 0 \text{ on } \Gamma,$$

where u is the given non-singular solution of the homogenized problem. Then it is shown that for sufficiently small ϵ , the hypotheses of the Newton-Kantorovich theorem can be verified if w_ϵ is used as the starting vector for the Newton method. Hence the Newton method will converge to a solution u_ϵ which will be unique in a neighbourhood of w_ϵ and of u . Thus $\{u_\epsilon\}$ will converge to u owing to the property of local uniqueness.

In § 2, the important hypotheses are made and some preliminary results are recalled. In § 3, the existence of solutions to the various problems is proved and the homogenized problem is identified. In § 4, the existence of a sequence $\{u_\epsilon\}$ converging to a given solution u is proved. § 5 is reserved for conclusions and various comments.

2. Preliminaries

The following hypotheses are made on the coefficients a_{ij}^ϵ :

(H1) There exists a constant $M > 0$ independent of ϵ such that

$$(2.1) \quad |a_{ij}^\epsilon(x)| \leq M \text{ a.e., } 1 \leq i, j \leq n.$$

(H2) There exists a constant $\alpha_0 > 0$ such that

$$(2.2) \quad \forall \xi = (\xi_i) \in \mathbb{R}^n, a_{ij}(x) \xi_i \xi_j \geq \alpha_0 \xi_i \xi_i \text{ a.e.}$$

Definition 2.1. An operator A of the form,

$$(2.3) \quad A = - \frac{\partial}{\partial x_i} \left(a_{ij}(x) \frac{\partial}{\partial x_j} \right),$$

is said to be the *homogenized operator*, w.r.t. the family $\{A^\epsilon\}$ if the following holds:

Let $u_\epsilon \rightarrow u$ in $H_0^1(\Omega)$ weakly and $A^\epsilon u_\epsilon \rightarrow g$ in $H^{-1}(\Omega)$ strongly. Then

$$\xi_i^\epsilon = a_{ij}^\epsilon(x) \frac{\partial u_\epsilon}{\partial x_j} \rightarrow a_{ij}(x) \frac{\partial u}{\partial x_j} = \xi_i,$$

in $L^2(\Omega)$ weakly, $1 \leq i \leq n$.

This definition is due to Tartar [8]. As a consequence, given $\{f_\epsilon\}$ in $H^{-1}(\Omega)$ converging strongly to f , and $\{u_\epsilon\}$ the unique solutions of the problems

$$(2.4) \quad \forall v \in H_0^1(\Omega), \quad a_\epsilon(u_\epsilon, v) = \langle f, v \rangle (H^{-1}, H_0^1)$$

where for $w, v \in H_0^1(\Omega)$,

$$(2.5) \quad a_\epsilon(w, v) = \int_{\Omega} a_{ij}^\epsilon(x) \frac{\partial w}{\partial x_j} \frac{\partial v}{\partial x_i} dx,$$

then, $u_\epsilon \rightarrow u$ in $H_0^1(\Omega)$ weakly [and hence in $L^2(\Omega)$ strongly] to the unique solution, u , of the homogenized problem,

$$(2.6) \quad \forall v \in H_0^1(\Omega), \quad a(u, v) = \langle f, v \rangle (H^{-1}, H_0^1)$$

where for $w, v \in H_0^1(\Omega)$,

$$(2.7) \quad a(w, v) = \int_{\Omega} a_{ij}(x) \frac{\partial w}{\partial x_j} \frac{\partial v}{\partial x_i} dx.$$

The existence of such an operator is known (cf. Tartar [7]). The operator is also elliptic with the same constant of ellipticity. The coefficients are bounded. (If they are symmetric, the same bound works.) In case of the a_{ij}^ϵ having a periodic structure the operator A can be explicitly written (cf. Bensoussan *et al* [1]).

Let $f: \mathbb{R} \rightarrow \mathbb{R}$ be a function which verifies the following hypothesis :

(H3) f is a C^1 function such that both f and f' are Lipschitz continuous with constants K_0 and K_1 respectively. Further, it is assumed that $K_0 < \alpha$ the ellipticity constant of the bilinear forms $a_\epsilon(\cdot, \cdot)$ and $a(\cdot, \cdot)$.

Proposition 2.1. Suppose that the hypothesis (H3) is verified. Then if $u \in L^p(\Omega)$ for some $p \geq 1$, $f(u)$ and $f'(u)$ are also in $L^p(\Omega)$. If $u_\epsilon \rightarrow u$ in $L^p(\Omega)$, then $f(u_\epsilon) \rightarrow f(u)$ and $f'(u_\epsilon) \rightarrow f'(u)$ in $L^p(\Omega)$.

Proof.

$$\int_{\Omega} |f(u(x)) - f(0)|^p dx \leq K_0^p \int_{\Omega} |u(x)|^p dx < +\infty.$$

Since constant functions are in $L^p(\Omega)$ for Ω bounded it follows that $f(u) \in L^p(\Omega)$. Similarly

$$\int_{\Omega} |f(u_\epsilon(x)) - f(u(x))|^p dx \leq K_0^p \int_{\Omega} |u_\epsilon(x) - u(x)|^p dx,$$

which converges to zero. Thus the assertions proved for $f(u)$. The proof for $f'(u)$ is identical.

The problem (1.2)–(1.3) can now be written as follows in its weak form :

(P_ϵ) To find $u_\epsilon \in H_0^1(\Omega)$ such that

$$(2.8) \quad \forall v \in H_0^1(\Omega), \quad a_\epsilon(u_\epsilon, v) = \int_{\Omega} f(u_\epsilon) \cdot v dx.$$

One can define the nonlinear operator

$T_\epsilon: H_0^1(\Omega) \rightarrow H_0^1(\Omega)$ as follows;

$$(2.9) \quad \forall v \in H_0^1(\Omega), \quad a_\epsilon(T_\epsilon(u), v) = \int_{\Omega} f(u) \cdot v \, dx.$$

Then the problem (P_ϵ) reduces to finding fixed points of T_ϵ or the zeros of F_ϵ where

$$(2.10) \quad F_\epsilon = I - T_\epsilon.$$

Remark 2.1. Since the injection $H_0^1(\Omega) \rightarrow L^2(\Omega)$ is compact, it follows that T_ϵ is compact.

Analogously, the operators T and $F(=I - T)$, associated to the homogenized operator A , can be defined.

Finally, an abstract theorem relating to solution of nonlinear equations in Banach spaces will now be recalled.

Theorem 2.1. Let X and Y be Banach spaces and $F: D \subset X \rightarrow Y$. Suppose that on an open convex set $D_0 \subset D$, F is Fréchet differentiable and

$$(2.11) \quad \forall x, y \in D_0, \quad \|F'(x) - F'(y)\| \leq K \|x - y\|.$$

For some $x_0 \in D_0$, assume that $\Gamma_0 = (F'(x_0))^{-1}$ is defined on all of Y and that $h \equiv \beta K \eta \leq \frac{1}{2}$ where $\|\Gamma_0\| \leq \beta$, and $\|\Gamma_0 F x_0\| \leq \eta$. Set

$$(2.12) \quad t^* = \frac{1}{\beta K} (1 - \sqrt{1 - 2h}), \quad t^{**} = \frac{1}{\beta K} (1 + \sqrt{1 - 2h}).$$

Suppose that $S \equiv \{x \mid \|x - x_0\| \leq t^*\} \subset D_0$.

Then the Newton iterates $\{x_k\}$ given by

$$(2.13) \quad x_{k+1} = x_k - (F'(x_k))^{-1} F(x_k), \quad k = 0, 1, \dots$$

are all well defined, lie in S , and converge to a solution x^* of $F(x) = 0$, which is unique in $D_0 \cap \{x \mid \|x - x_0\| < t^{**}\}$. Moreover, if $h < \frac{1}{2}$ the order of convergence is at least quadratic.

Known as the Newton-Kantorovich Theorem, the above was proved by Kantorovich (cf. [2], [3]). Ortega [6] has given a simpler proof to the same. This theorem will be used to prove the main result in § 4.

3. Existence and homogenization results

Throughout this section we assume that the hypotheses (H1)–(H3) are verified.

Proposition 3.1. There exists a constant $C > 0$, independent of ϵ , such that

$$(3.1) \quad \|u_\epsilon\|_{1, \Omega} \leq C.$$

Proof. Set $v = u_\epsilon$ in (2.8). Then

$$\begin{aligned} a \|u_\epsilon\|_{1, \Omega}^2 &\leq a_\epsilon(u_\epsilon, u_\epsilon) = \int_{\Omega} f(u_\epsilon) u_\epsilon \, dx \\ &= \int_{\Omega} (f(u_\epsilon) - f(a)) u_\epsilon \, dx + \int_{\Omega} f(a) u_\epsilon \, dx, \quad f(a) = 0 \\ &\leq K_0 \int_{\Omega} |u_\epsilon - a| |u_\epsilon| \, dx + C_1 |u_\epsilon|_{0, \Omega} \\ &\leq K_0 \|u_\epsilon\|_{0, \Omega}^2 + C_2 |u_\epsilon|_{0, \Omega}^2, \end{aligned}$$

or

$$\|u_\epsilon\|_{1,\Omega}^2 \leq \frac{C_2}{\alpha - K_0} \|u_\epsilon\|_{0,\Omega} \leq \frac{C_2}{\alpha - K_0} \|u_\epsilon\|_{1,\Omega},$$

which proves (3.1).

Remark 3.1. In the above proof, the hypothesis $K_0 < \alpha$ has been used.

Theorem 3.1. There exists at least one solution, to the problem (P_ϵ) .

Proof. First of all it was remarked in the previous section that T_ϵ is compact. Further if $0 < \lambda < 1$ it can be proved that any solution of the problem

$$(3.2) \quad z_\epsilon = \lambda T_\epsilon z_\epsilon,$$

is such that

$$(3.3) \quad \|z_\epsilon\|_{1,\Omega} \leq C,$$

where $C > 0$ is as in Proposition 3.1. (The proof of this fact is identical to that of the above-mentioned proposition.) Thus by Schaeffer's Theorem T_ϵ admits at least one fixed point in $B(0; C)$, the ball centre 0 and radius C in $H_0^1(\Omega)$.

Theorem 3.2. There is at least one solution to the homogenized problem $(P) : u = Tu$

Proof. The proof is identical to that of the preceding theorem.

Theorem 3.3. Let u_ϵ be a solution to (P_ϵ) . Then there exists a subsequence (again indexed by ϵ) such that $u_\epsilon \rightarrow u$ in $H_0^1(\Omega)$ weakly, u being a solution to the homogenized problem (P) .

Proof. By Proposition 3.1, the u_ϵ are all bounded in $H_0^1(\Omega)$. Thus we can extract a weakly convergent subsequence with limit, say, u . Then $u_\epsilon \rightarrow u$ in $L^2(\Omega)$ strongly. Hence by proposition 2.1, $f(u_\epsilon) \rightarrow f(u)$ in $L^2(\Omega)$ strongly and thus in $H^{-1}(\Omega)$ strongly. Now by the definition of the homogenized operator (cf. definition 2.1) we have

$$(3.4) \quad \forall v \in H_0^1(\Omega), \quad a(u, v) = \int_{\Omega} f(u) v \, dx,$$

which proves the theorem.

The idea of using Schaeffer's Theorem is quite classical in the existence theory of nonlinear equations (cf. Rabinowitz [7]).

4. The main theorem

Under the hypotheses (H1)–(H3), we already know that the problems (P) and (P_ϵ) admit at least one solution each. We will prove now that if $n \leq 5$, and u a non-singular solution of (P) , among the solution of (P_ϵ) there exists a solution u_ϵ which is non singular, tending to u . This is an important step toward the study of the stability of homogenization for the associated evolution equation.

Proposition 4.1. Let $n \leq 6$. Then T_ϵ (resp. T) is Fréchet differentiable on all of V . The family $\{T_\epsilon, T\}$ is uniformly Lipschitz continuous.

Proof. Since $n \leq 6$, the following inclusion holds :

$$H_0^1(\Omega) \rightarrow L^3(\Omega).$$

Then for $u, v, w \in H_0^1(\Omega)$, the integral

$$\int f'(u) v w \, dx,$$

is well defined and is continuous w.r.t. v and w . It is easy to check that the Fréchet derivative of T_ϵ at u is defined by

$$(4.1) \quad \forall w \in H_0^1(\Omega), a_\epsilon(T'_\epsilon(u) v, w) = \int_\Omega f'(u) v w \, dx,$$

for any $v \in H_1^0(\Omega)$. Now, let $u_1, u_2 \in H_1^0(\Omega)$. Set

$$(4.2) \quad z_i^\epsilon = T'_\epsilon(u_i) v, \quad i = 1, 2,$$

where $v \in H_1^0(\Omega)$. Then

$$\begin{aligned} a \|z_1^\epsilon - z_2^\epsilon\|_{1,\Omega}^2 &\leq a_\epsilon(z_1^\epsilon - z_2^\epsilon, z_1^\epsilon - z_2^\epsilon) \\ &= \int_\Omega (f'(u_1) - f'(u_2)) (z_1^\epsilon - z_2^\epsilon) v \, dx \\ &\leq K_1 \int |u_1 - u_2| |z_1^\epsilon - z_2^\epsilon| |v| \, dx \\ &\leq K_1 |u_1 - u_2|_{0,3,\Omega} |z_1^\epsilon - z_2^\epsilon|_{0,3,\Omega} |v|_{0,3,\Omega} \\ &\leq C K_1 \|u_1 - u_2\|_{1,\Omega} \|z_1^\epsilon - z_2^\epsilon\|_{1,\Omega} \|v\|_{1,\Omega}, \end{aligned}$$

$$\text{or} \quad \|(T'_\epsilon(u_1) - T'_\epsilon(u_2)) v\|_{1,\Omega} \leq K \|u_1 - u_2\|_{1,\Omega} \|v\|_{1,\Omega},$$

which gives

$$(4.3) \quad \|T'_\epsilon(u_1) - T'_\epsilon(u_2)\|_{1,\Omega} \leq K \|u_1 - u_2\|_{1,\Omega}$$

K being independent of ϵ . The proof for T is identical.

Proposition 4.2. Let $n \leq 5$. Then for any $u \in H_0^1(\Omega)$, $T'_\epsilon(u)$ (resp. $T'(u)$) is compact.

Proof. For $n \leq 5$, the injection

$$(4.4) \quad H_0^1(\Omega) \rightarrow L^3(\Omega)$$

is compact. The result is a direct consequence of this fact.

Remark 4.1. If $n = 6$, then the inclusion (4.4) is not compact and the above result does not hold.

Henceforth, it will be assumed that u is a given solution of the homogenized problem (P). Let, for each ϵ , $w_\epsilon \in H_0^1(\Omega)$ be defined as the unique solution of the following linear problem.

$$(4.5) \quad \forall v \in H_0^1(\Omega), a_\epsilon(w_\epsilon, v) = \int_\Omega f(u) v \, dx.$$

From the theory of homogenization of linear problems, it follows that $w_\epsilon \rightarrow u$ in $H_0^1(\Omega)$ weakly [and in $L^2(\Omega)$ strongly].

Proposition 4.3. Let $n \leq 5$. Let u be a non-singular solution of problem (P), i.e. $F'(u) = I - T'(u)$ is invertible. Then for sufficiently small ϵ , $F'_\epsilon(w_\epsilon) =$

$I - T'_\epsilon(w_\epsilon)$ is also invertible. Further, there exists a $\beta > 0$, independent of ϵ , such that

$$(4.6) \quad \|(I - T'_\epsilon(w_\epsilon))^{-1}\| \leq \beta.$$

Proof. Assume the first assertion false. Then there exists a sequence $\epsilon_n \rightarrow 0$ such that

$$I - T'_{\epsilon_n}(w_{\epsilon_n}),$$

is singular. (Henceforth, purely as a matter of convenience, the index n will be suppressed.) Since $T'_\epsilon(w_\epsilon)$ is compact, there exists a non-zero function $z_\epsilon \in H_0^1(\Omega)$ such that

$$(4.7) \quad z_\epsilon = T'_\epsilon(w_\epsilon) z_\epsilon.$$

It can be further assumed that

$$(4.8) \quad \|z_\epsilon\|_{1,\Omega} = 1.$$

Now (4.7) can be rewritten as

$$(4.9) \quad \forall w \in H_0^1(\Omega), \quad a_\epsilon(z_\epsilon, w) = \int_\Omega f'(w_\epsilon) z_\epsilon w \, dx.$$

Now for a subsequence (again indexed by ϵ) $z_\epsilon \rightarrow z$ and $w_\epsilon \rightarrow u$ in $H_0^1(\Omega)$ weakly. Consider, for any $w \in H_0^1(\Omega)$,

$$\begin{aligned} & \left| \int_\Omega (f'(w_\epsilon) z_\epsilon - f'(u) z) w \, dx \right| \\ & \leq \int_\Omega |f'(w_\epsilon)| |z_\epsilon - z| |w| \, dx + \int_\Omega |f'(w_\epsilon) - f'(u)| |z| |w| \, dx \\ & \leq C(|z_\epsilon - z|_{0,3,\Omega} + |f'(w_\epsilon) - f'(u)|_{0,3,\Omega}) \|w\|_{1,\Omega}. \end{aligned}$$

Since $z_\epsilon \rightarrow z$ in $L^3(\Omega)$ strongly ($n \leq 5$) and $f'(w_\epsilon) \rightarrow f'(u)$ in $L^3(\Omega)$ strongly [cf. Proposition (2.11)] it follows that

$$f'(w_\epsilon) z_\epsilon \rightarrow f'(u) z,$$

in $H^{-1}(\Omega)$ strongly. Then by the definition of the homogenized operator,

$$(4.10) \quad \forall w \in H_0^1(\Omega), \quad a(z, w) = \int_\Omega f'(u) z w \, dx,$$

or, equivalently,

$$(4.11) \quad z = T'(u) z.$$

By assumption, this is possible only if $z = 0$. Hence $z_\epsilon \rightarrow 0$ in $H_0^1(\Omega)$ weakly ($L^3(\Omega)$ strongly).

Now, choosing $w = z_\epsilon$ in (4.9) and taking into account (4.8), it follows that

$$0 < \alpha \leq \int_\Omega f'(w_\epsilon) z_\epsilon z_\epsilon \, dx \leq |f'(w_\epsilon)|_{0,3,\Omega} |z_\epsilon|_{0,3,\Omega},$$

which converges to zero, thus giving a contradiction. Hence the first assertion is true.

Assume now that (4.6) is false. Then there exists a sequence (denoted, as usual, by ϵ itself) y_ϵ and a sequence v_ϵ in $H_0^1(\Omega)$ such that

$$(4.12) \quad \begin{cases} v_\epsilon = (I - T'_\epsilon(w_\epsilon))^{-1} y_\epsilon, \\ \|v_\epsilon\|_{1, \Omega} = 1, \|y_\epsilon\|_{1, \Omega} \rightarrow 0. \end{cases}$$

Set,

$$(4.13) \quad z_\epsilon = T'_\epsilon(w_\epsilon) v_\epsilon.$$

Then

$$(4.14) \quad y_\epsilon = v_\epsilon - z_\epsilon.$$

Clearly, $\|z_\epsilon\|_{1, \Omega} \leq C$ and hence (for a subsequence) $v_\epsilon \rightarrow v$, $z_\epsilon \rightarrow z$, say, in $H_0^1(\Omega)$ weakly. Since $y_\epsilon \rightarrow 0$ in $H_0^1(\Omega)$ strongly it follows that $z = v$. As before it can be shown that

$$f'(w_\epsilon) v_\epsilon \rightarrow f'(u) v,$$

in $H^{-1}(\Omega)$ strongly. Passing to the limit,

$$(4.15) \quad \forall w \in H_0^1(\Omega), a(z, w) = \int_{\Omega} f'(u) z w \, dx,$$

which again implies $z = 0 = v$. Once again,

$$(4.16) \quad \alpha \|z_\epsilon\|_{1, \Omega}^2 \leq a_\epsilon(z_\epsilon, z_\epsilon) = \int_{\Omega} f'(w_\epsilon) v_\epsilon z_\epsilon \, dx.$$

The integral converges to zero since $v_\epsilon \rightarrow 0$, $z_\epsilon \rightarrow 0$ in $L^3(\Omega)$ strongly. Hence it follows from (4.16) that $z_\epsilon \rightarrow 0$ in $H_0^1(\Omega)$ strongly and hence from (4.14) that $v_\epsilon \rightarrow 0$ in $H_0^1(\Omega)$ strongly. But this contradicts the fact that $\|v_\epsilon\|_{1, \Omega} = 1$. Thus, (4.6) is established.

Proposition 4.4.

$$\lim_{\epsilon \rightarrow 0} \|F_\epsilon(w_\epsilon)\|_{1, \Omega} = 0.$$

Proof. Set

$$(4.17) \quad y_\epsilon = T_\epsilon w_\epsilon.$$

$$\text{Then } \|F_\epsilon(w_\epsilon)\|_{1, \Omega} = \|w_\epsilon - y_\epsilon\|_{1, \Omega}.$$

Now by definition,

$$(4.18) \quad \forall w \in H_0^1(\Omega), a_\epsilon(w_\epsilon - y_\epsilon, w) = \int_{\Omega} (f(u) - f(w_\epsilon)) w \, dx.$$

Thus, it follows that

$$\alpha \|w_\epsilon - y_\epsilon\|_{1, \Omega}^2 \leq K_0 \|u - w_\epsilon\|_{0, \Omega} \|w_\epsilon - y_\epsilon\|_{0, \Omega},$$

or, for some $C > 0$, independent of ϵ ,

$$(4.19) \quad \|w_\epsilon - y_\epsilon\|_{1, \Omega} \leq C \|u - w_\epsilon\|_{0, \Omega},$$

which converges to zero, thus proving the result.

The main theorem can now be stated and proved.

Theorem 4.1. Let the hypotheses (H1-H3) be true. Let $n \leq 5$. If u is a given non-singular solution of the homogenized problem (P), then for sufficiently small ϵ , the problem (P_ϵ) admits a solution u_ϵ such that $u_\epsilon \rightarrow u$ in $H_0^1(\Omega)$ weakly and $L^2(\Omega)$ strongly. Further, the solution u_ϵ is also non-singular.

Proof. For the first assertion, it will be shown that all the hypotheses of the Newton-Kantorovich theorem are verified.

Step 1. First of all for every ϵ , F'_ϵ exists on all of $H_0^1(\Omega)$ and is Lipschitz continuous with a constant K , independent of ϵ (Proposition 4.1).

Step 2. Consider w_ϵ defined by (4.5). Then for sufficiently small ϵ , $\Gamma_{0,\epsilon} = (F'_\epsilon(w_\epsilon))^{-1}$ exists and $\|\Gamma_{0,\epsilon}\| \leq \beta$, independent of ϵ (Proposition 4.3).

Step 3. Finally set

$$(4.20) \quad \eta_\epsilon = C\beta \|u - w_\epsilon\|_{0,\Omega},$$

where C is as in (4.19). Then

$$\|\Gamma_{0,\epsilon} F'_\epsilon(w_\epsilon)\| \leq \eta_\epsilon,$$

and $\eta_\epsilon \rightarrow 0$ as $\epsilon \rightarrow 0$. Hence for sufficiently small ϵ ,

$$(4.21) \quad h_\epsilon = K\beta \eta_\epsilon < 1/2.$$

In fact, $h_\epsilon \rightarrow 0$ as $\epsilon \rightarrow 0$. Thus the Newton iterates

$$(4.22) \quad \begin{cases} u_\epsilon^{k+1} = u_\epsilon^k - (F'_\epsilon(u_\epsilon^k))^{-1} F_\epsilon(u_\epsilon^k) \\ u_\epsilon^0 = w_\epsilon \end{cases}$$

are all well-defined, lie in the ball

$$(4.23) \quad S_\epsilon = \{v \mid \|v - w_\epsilon\|_{1,\Omega} \leq t_\epsilon^*\},$$

with

$$(4.24) \quad t_\epsilon^* = 1/\beta K (1 - \sqrt{1 - 2h_\epsilon}),$$

and converge to a solution u_ϵ of (P_ϵ) which also lies in S_ϵ , by the Newton-Kantorovich theorem (cf. Theorem 2.1).

Step 4. Now by Theorem 3.3, given a subsequence of $\{u_\epsilon\}$, a further subsequence can be extracted converging to a solution, say, $\hat{u}$ of (P). But since $h_\epsilon \rightarrow 0$, it follows that $t_\epsilon^* \rightarrow 0$ and hence $\|u_\epsilon - w_\epsilon\|_{1,\Omega} \rightarrow 0$. Since both u_ϵ and w_ϵ converge weakly, the limits must coincide and hence $\hat{u} = u$. Thus irrespective of the subsequence, the limit is always u and so it can be deduced the entire family u_ϵ converges to u weakly in $H_0^1(\Omega)$.

Step 5. To prove the non-singularity of u_ϵ , the invertibility of $I - T'_\epsilon(u_\epsilon)$ must be proved. This is done exactly as in Proposition (4.3).

Thus the theorem is completely proved.

Remark 4.2. It may be wondered that u_0 could be used as a starting point for the Newton iterates for each ϵ instead of w_ϵ , thus avoiding several estimates. But then $F_\epsilon(u_0) = u_0 - w_\epsilon$ which does not converge to zero in $H_0^1(\Omega)$. Thus $h_\epsilon < 1/2$ cannot be guaranteed. Thus u_ϵ converges to u only weakly while w_ϵ is always sufficiently close to u_ϵ to guarantee the convergence of the Newton method

5. Conclusion

It has been proved that non-singular solutions of homogenized problem are limits of non-singular solutions of the given family of problems, (P_ϵ) . The hypotheses on the coefficients of the differential operator and on the given function f are fairly minimal.

The hypotheses (H1) and (H2) are standard ones used in the linear theory of homogenization. The Lipschitz continuity of f and f' laid down in hypothesis (H3) was used to prove the Fréchet derivability of T_ϵ and the Lipschitz continuity of T'_ϵ , an essential condition in the Newton-Kantorovich Theorem. It seems possible that, working with classical solutions (cf. Mythily [5]), the local Lipschitz continuity of f and f' may suffice. The only simplifying assumption is $K_0 < a$, which has been used only to prove the existence of a solution *via a priori* estimates. Indeed this condition can be removed and replaced by any other which guarantees the existence of a non-singular solution of the homogenized problem alone. Then the method of this paper can be used to prove (without new hypotheses) the existence of a solution of (P_ϵ) . This solution will be non-singular and will converge to u .

The principal drawback seems to be the restriction on the dimension, viz., $n \leq 5$. In the case $n = 6$, the theorem can be established with the additional hypotheses that f' is bounded. Then $f'(u)v \in L^2(\Omega)$ for $u, v \in H_0^1(\Omega)$ and the compactness of the operators $T'_\epsilon(u)$ and $T''(u)$ will follow.

The case $n > 6$ seems to be open. The main difficulty in this case is the differentiability of T_ϵ . It is not obvious [even for $f'(x) = x$] that $f'(u)v \in H^{-1}(\Omega)$. Of course, it will be so if f' is assumed bounded. Even then while $T'(u)$ can be formally defined by (4.1), it cannot be proved that it is indeed the Fréchet derivative of T_ϵ at u . This is because in assuming f' to be bounded, while $f'(u)v \in L^2(\Omega) \rightarrow H^{-1}(\Omega)$, $f'(u)$ does not play any role in the estimates. Again for the same reasons T'_ϵ cannot be proved to be Lipschitz continuous.

References

- [1] Bensoussan A, Lions J L and Papanicolaou G 1978 *Asymptotic methods in periodic structure* (Amsterdam : North Holland)
- [2] Kantorovich L V 1948 *Usp. Mat. Nauk.* 3 89
- [3] Kantorovich L V and Akilov G P 1959 *Functional analysis in normed spaces* (Engl. Trans.) (New York : Pergamon)
- [4] Kesavan S 1979 *Appl. Math. Optim.* 5 153
- [5] Mythily R 1981 On a nonlinear elliptic equation with periodic coefficients (to appear)
- [6] Ortega J M 1968 *Am. Math. Monthly* 75 658
- [7] Rabinowitz P H 1973 *Theorie du Degré topologique et applications a des problèmes aux limites non-linéaires* (Paris : Cours 6)
- [8] Tartar L 1981 Cours peccot, Collège de France, Paris (to appear)

Homogenization of eigenvalue problems in perforated domains

M VANNINATHAN

TIFR Centre, Indian Institute of Science, Bangalore 560 012, India

MS received 9 December 1980 ; revised 3 January 1981

Abstract. In this paper, we treat some eigenvalue problems in periodically perforated domains and study the asymptotic behaviour of the eigenvalues and the eigenvectors when the number of holes in the domain increases to infinity. Using the method of asymptotic expansion, we give explicit formula for the homogenized coefficients and expansion for eigenvalues and eigenvectors. If we denote by ϵ the size of each hole in the domain, then we obtain the following asymptotic expansion for the eigenvalues :

$$\text{Dirichlet : } \lambda_{\epsilon} = \epsilon^{-2} \lambda + \lambda_0 + O(\epsilon),$$

$$\text{Stekloff : } \lambda_{\epsilon} = \epsilon \lambda_1 + O(\epsilon^2),$$

$$\text{Neumann : } \lambda_{\epsilon} = \lambda_0 + \epsilon \lambda_1 + O(\epsilon^2).$$

Using the method of energy, we prove a theorem of convergence in each case considered here. We briefly study correctors in the case of Neumann eigenvalue problem.

Keywords. Homogenization ; correctors ; eigenvalues ; eigenvectors.

1. Introduction

The theory of homogenization has been developed by many authors in recent years. For a historic introduction and for a complete bibliography of the subject, the reader is referred to the book of Bensoussan *et al* [3]. The method of asymptotic development introduced in this book can also be applied to problems in a periodically perforated domain. For the treatment of homogenization problems in such domains, the reader is referred to the works of Lions [13], Duvaut [9], Cioranescu [7], Cioranescu and Saint Jean Paulin [8].

The study of such problems is important from theoretical as well as numerical point of view. Because of the complicated structure of the perforated domains, any kind of calculation is difficult to perform. For example, if we treat the Dirichlet problem, we have to impose the boundary condition on the boundary of the holes which are many in number. So, we would like to "approximate" the given problem by a "homogenized" problem on the domain without holes,

By the method of asymptotic development, a problem on a periodically perforated domain is reduced to solving problems in the "basic cell" and in the domain without holes.

This paper is divided into three parts :

Part A : Dirichlet eigenvalue problem

Part B : Stekloff eigenvalue problem

Part C : Neumann eigenvalue problem.

Our aim is to describe the asymptotic behaviour of the various eigenvalues when the number of holes in the domain increases to infinity. In each case we explicitly write down the "homogenized operator" with the help of the method of asymptotic development and prove a homogenization theorem using the energy method introduced by Tartar [14] and prolongation operators of Cioranescu and Saint Jean Paulin [8].

We treat here the case of Laplacian operator. But one can extend the results to the case of elliptic, self-adjoint operators with periodic coefficients of the form

$$A^\epsilon \equiv - \sum_{i,j} \frac{\partial}{\partial x_i} \left(a_{ij}(x/\epsilon) \frac{\partial}{\partial x_j} \right).$$

The eigenvalue problem corresponding to A^ϵ in a fixed domain has been studied by Kesavan [12]. The results of this paper were announced in Vanninathan [15], [16].

2. Notations and hypothesis

First, we consider a reference cell :

$$(2.1) \quad Y = \prod_{i=1}^N (0, y_i^0) \subset \mathbf{R}^N, y_i^0 > 0.$$

Let τ_i ($i = 1, 2, \dots, M$) be connected bounded open subsets of $\mathbf{R}^N$ with sufficiently smooth boundaries and which lie locally on one side of the boundary. Then the holes in Y are $\bar{\tau}_i \cap Y$ ($i = 1, 2, \dots, M$) and their union is denoted by T :

$$(2.2) \quad T = \bigcup_{i=1}^M (\bar{\tau}_i \cap Y).$$

Let

$$(2.3) \quad Y^* = Y - T.$$

Let S denote the boundary of T in Y^* . For details see figure 1.

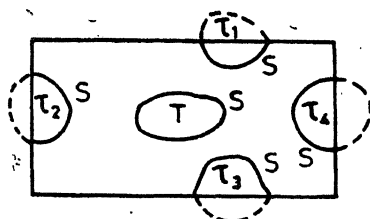


Figure 1

We assume that

(H1) Y^* is connected.

For a function f given on Y , we define the average of f on Y :

$$(2.4) \quad m_Y(f) = \frac{1}{|Y|} \int_Y f(y) dy,$$

where $|Y|$ denotes the Lebesgue measure of Y . Let ϵ be a small positive parameter which goes to zero. We denote by $f^\epsilon(x)$ or $f(x/\epsilon)$, the function defined in $\mathbb{R}^N$ in a periodic fashion with period ϵy_i^0 in the direction x_i .

Let us define now, "the holes" in $\mathbb{R}^N$ corresponding to ϵ , starting from those in Y . For that we introduce the function χ as follows:

$$(2.5) \quad \chi(y) = \begin{cases} 1 & \text{if } y \in Y^*, \\ 0 & \text{if } y \in T. \end{cases}$$

We consider also the characteristic function of T :

$$(2.6) \quad \chi_T(y) = \begin{cases} 1 & \text{if } y \in T, \\ 0 & \text{if } y \in Y^*. \end{cases}$$

Put

$$(2.7) \quad \theta = m_Y(\chi) = \frac{|Y^*|}{|Y|}.$$

The "holes" in $\mathbb{R}^N$ are the connected components of the set

$$\{x \in \mathbb{R}^N / \chi^\epsilon(x) = 0\}.$$

Finally, let us be given a bounded connected open subset Ω of $\mathbb{R}^N$ whose boundary is sufficiently smooth and Ω lies locally on one side of its boundary. The "holes" in Ω are then defined by

(2.8) T_ϵ = connected components in Ω of $\{x \in \Omega / \chi^\epsilon(x) = 0\}$. The perforated domain Ω_ϵ with which we work is

$$(2.9) \quad \Omega_\epsilon = \Omega - T_\epsilon.$$

Let S_ϵ denote the boundary of T_ϵ in Ω_ϵ .

We make the following restrictions on the geometry of Ω_ϵ .

(H2) Ω_ϵ is connected.

(H3) Each hole in T_ϵ has regular boundary.

In the problems we consider here, there is one more restriction on the geometry of Ω and the holes (cf. Cioranescu and Saint Jean Paulin [8]).

(H4) The holes T_ϵ do not meet $\partial\Omega$, the boundary of Ω .

We need, in fact, in Part A a stronger hypothesis. Given any hole T' in Y , we can as before construct the holes T'_ϵ in Ω periodically. Set

$$(2.10) \quad \Omega_{T'_\epsilon} = \Omega - T'_\epsilon.$$

With this notation, it is evident that

$$(2.11) \quad \Omega_\epsilon = \Omega_{T_\epsilon}.$$

The stronger hypothesis is the following:

(H5) $\begin{cases} \text{there exists a hole } T' \text{ in } Y \text{ such that } T \subset \subset T' \text{ and the holes } T'_\epsilon \text{ do} \\ \text{not intersect } \partial\Omega. \end{cases}$

Remark (2.1). The hypothesis (H4) is severe on the geometry of the domain Ω and the hole T . One such example of Ω is a finite union of cells homothetic to Y and with the hole T placed in the middle of Y .

Summation Convention. We adopt the usual summation convention with respect to the repeated indices.

Part A: Dirichlet eigenvalue problem

3. Problem to be treated

With the above notation, we consider the following eigenvalue problem:

$$(3.1) \quad \begin{cases} \text{Find } (u_\epsilon, \lambda_\epsilon) \in H_0^1(\Omega_\epsilon) \times \mathbf{R} \text{ such that} \\ -\Delta u_\epsilon = \lambda_\epsilon u_\epsilon \text{ in } \Omega_\epsilon, \\ (u_\epsilon, u_\epsilon)_\epsilon = 1, \end{cases}$$

where $(\cdot, \cdot)_\epsilon$ denotes the inner product in $L^2(\Omega_\epsilon)$ and $H_0^1(\Omega_\epsilon) = \{v \in L^2(\Omega_\epsilon); \partial v / \partial x_i \in L^2(\Omega_\epsilon) \text{ for } i = 1, 2, \dots, N \text{ and } v = 0 \text{ on } \partial\Omega\}$.

The variational formulation of this problem is the following :

$$(3.2) \quad \begin{cases} \text{Find } (u_\epsilon, \lambda_\epsilon) \in H_0^1(\Omega_\epsilon) \times \mathbf{R} \text{ such that} \\ a^\epsilon(u_\epsilon, v) = \lambda_\epsilon (u_\epsilon, v)_\epsilon \text{ for } v \in H_0^1(\Omega_\epsilon), \\ (u_\epsilon, u_\epsilon)_\epsilon = 1, \end{cases}$$

with the bilinear form $a^\epsilon(\cdot, \cdot)$ defined by

$$(3.3) \quad a^\epsilon(u, v) = \int_{\Omega_\epsilon} \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_i} dx.$$

According to spectral theory, there exist a sequence of eigenvalues $\{\lambda_\epsilon^l\}_{l=1}^\infty$ and a sequence of corresponding eigenvectors $\{u_\epsilon^l\}_{l=1}^\infty$ such that

$$(3.4) \quad \begin{cases} 0 < \lambda_\epsilon^1 < \lambda_\epsilon^2 \leq \lambda_\epsilon^3 \dots \rightarrow \infty, \\ \lambda_\epsilon^l \text{ is of finite multiplicity for each } l, \\ \text{and } \{u_\epsilon^l\}_{l=1}^\infty \text{ form an orthonormal basis in } L^2(\Omega_\epsilon). \end{cases}$$

We can characterise the eigenvalues λ_ϵ^l with the help of Rayleigh quotient (*cf.* Weinstein and Steinger [17]).

$$(3.5) \quad R_\epsilon(v) = \frac{a^\epsilon(v, v)}{(v, v)_\epsilon} \text{ for } v \in H_0^1(\Omega_\epsilon), v \neq 0.$$

The minimax principle for the eigenvalues states that

$$(3.6) \quad \begin{cases} \lambda_\epsilon^l = \min_{v \in S_l} \{ \max_{v \in S_l} R_\epsilon(v); S_l \subset H_0^1(\Omega_\epsilon), \dim S_l = l \}, \\ = \max_{v \in E_\epsilon(l)} R_\epsilon(v) \\ = \max \{ R_\epsilon(v); (v, u_\epsilon^i)_\epsilon = 0, i = 1, 2, \dots, l-1 \}, \end{cases}$$

where $E_\epsilon(l)$ is the subspace of $H_0^1(\Omega_\epsilon)$ spanned by $\{u_\epsilon^1, \dots, u_\epsilon^l\}$.

This part is devoted to the study of the behaviour of λ_ϵ^l and u_ϵ^l when $\epsilon \rightarrow 0$. We prove, in particular, that λ_ϵ^l is of order ϵ^{-2} and that $\{\lambda_\epsilon^l - \epsilon^{-2} \lambda\} \rightarrow l$ th eigenvalue of the "homogenized problem" where λ is the first eigenvalue in the cell

Y^* . As we will see, some weighted Sobolev spaces and their properties are used in this study.

4. Eigenvalue problem in the cell Y^*

Let us define the space

(4.1) $W_0 = \{v \in H^1(Y^*); v = 0 \text{ on } S \text{ and } v \text{ is } Y\text{-periodic, i.e., } v \text{ assumes same values on the opposite faces of } Y\}$
and the bilinear form

$$(4.2) \quad a(u, v) = \int_{Y^*} \frac{\partial u}{\partial y_i} \frac{\partial v}{\partial y_i} dy \text{ for } u, v \in H^1(Y^*).$$

The bilinear form being elliptic on W_0 , the following problem is well posed :

$$(4.3) \quad \begin{cases} \text{Find } (\phi, \lambda) \in W_0 \times \mathbf{R} \text{ such that} \\ a(\phi, v) = \lambda (\phi, v)_{Y^*} \text{ for } v \in W_0, \\ (\phi, \phi)_{Y^*} = 1, \end{cases}$$

where $(\cdot, \cdot)_{Y^*}$ denotes the scalar product in $L^2(Y^*)$.

In what follows, we consider only the first eigenvalue λ of the above problem. It is well known that λ is simple and the corresponding eigenvector ϕ has constant sign in Y^* . We choose the vector ϕ which is uniquely defined by (4.3) and

$$(4.4) \quad \phi > 0 \text{ in } Y^*.$$

Remark (4.1). We extend ϕ by zero in the interior of the holes T and we denote again by ϕ the extended function.

Remark (4.2). It follows from (4.3) that the function ϕ^ϵ defined periodically satisfies

$$(4.5) \quad \begin{cases} -\Delta \phi^\epsilon = \epsilon^{-2} \lambda \phi^\epsilon \text{ in } \Omega_\epsilon, \\ \phi^\epsilon = 0 \text{ on } S_\epsilon. \end{cases}$$

However, ϕ^ϵ is not zero on $\partial\Omega$.

5. Some weighted Sobolev spaces

We will see later that consideration of some weighted Sobolev spaces is very important in the study of the present problem. There is a vast literature on this subject: see for example Baouendi [1], Baouendi and Goulaouic [2], Geymont and Grisvard [10], Goudjo [11]. In this section, we define some weighted Sobolev spaces and state some of their properties which will be needed later.

We consider the following spaces with weights ϕ^ϵ and ϕ (ϕ being defined in § 4):

$$(5.1) \quad V_\epsilon = \left\{ v \in D'(\Omega_\epsilon); \phi^\epsilon v \in L^2(\Omega_\epsilon), \phi^\epsilon \frac{\partial v}{\partial x_i} \in L^2(\Omega_\epsilon) \text{ for } \right. \\ \left. i = 1, 2, \dots, N \text{ and } v = 0 \text{ on } \partial\Omega \right\},$$

$$(5.2) \quad V = \left\{ v \in D'(Y^*); \phi v \in L^2(Y^*), \phi \frac{\partial v}{\partial y_i} \in L^2(Y^*) \text{ for } j = 1, 2, \dots, N \text{ and } v \text{ is } Y\text{-periodic} \right\}.$$

where, as usual,

$D'(\Omega)$ = space of distributions on Ω , and these spaces are provided with the following norms :

$$\|v\|_\epsilon = \left[\|\phi^\epsilon v\|_{L^2(\Omega_\epsilon)}^2 + \sum_{i=1}^N \left\| \phi^\epsilon \frac{\partial v}{\partial x_i} \right\|_{L^2(\Omega_\epsilon)}^2 \right]^{1/2},$$

$$\|v\|_v = \left[\|\phi v\|_{L^2(Y^*)}^2 + \sum_{i=1}^N \left\| \phi \frac{\partial v}{\partial y_i} \right\|_{L^2(Y^*)}^2 \right]^{1/2}.$$

Some of their properties are given in the following :

Proposition (5.1)

(1) The space $D(\Omega_\epsilon)$ of infinitely differentiable functions on Ω_ϵ with compact support in Ω_ϵ is dense in V_ϵ .

(2) We have a continuous inclusion

$$V_\epsilon \rightarrow L^2(\Omega_\epsilon)$$

and the continuity constant of this inclusion does not depend on ϵ : there exists a constant $c > 0$ independent of ϵ such that

$$(5.5) \quad \|v\|_{L^2(\Omega_\epsilon)} \leq c \|v\|_\epsilon \quad \forall v \in V_\epsilon.$$

(3) The map $v \rightarrow \phi^\epsilon v$ defines an isomorphism of V_ϵ onto $H_0^1(\Omega_\epsilon)$.

(4) The inclusion $V_\epsilon \rightarrow L^2(\phi^\epsilon)$ is compact where

$$(5.6) \quad L^2(\phi^\epsilon) = \{v \in D'(\Omega_\epsilon); \phi^\epsilon v \in L^2(\Omega_\epsilon)\}.$$

$$(5) \quad v \rightarrow \left[\sum_{i=1}^N \left\| \phi^\epsilon \frac{\partial v}{\partial x_i} \right\|_{L^2(\Omega_\epsilon)}^2 \right]^{1/2}$$

defines a norm on V_ϵ equivalent to the norm $\| \cdot \|_\epsilon$.

Proof. All these properties are proved in the works cited above. The only thing which is new is the inequality (5.5) with c independent of ϵ . But one can prove this, without much difficulty, from the continuity of the inclusion

$$(5.7) \quad W(\phi) \rightarrow L^2(Y^*),$$

where the space $W(\phi)$ is defined by

$$(5.8) \quad W(\phi) = \left\{ v \in D'(Y^*); \phi v \in L^2(Y^*), \phi \frac{\partial v}{\partial y_i} \in L^2(Y^*) \text{ for } i = 1, 2, \dots, N. \right\}$$

Proposition (5.2)

(1) The space of functions $v \in C^\infty(\bar{Y}^*) \cap V$ which vanish in a neighbourhood of S are dense in V .

(2) One has the continuous inclusion

$$V \rightarrow L^2(Y^*).$$

(3) The map $v \rightarrow \phi v$ defines an isomorphism of V onto W_0 .

(4) The inclusion

$$(5.9) \quad V \rightarrow L^2(\phi)$$

is compact where we define

$$(5.10) \quad L^2(\phi) = \{v \in D'(Y^*) ; \phi v \in L^2(Y^*)\}.$$

$$(5) \quad v \rightarrow \left[\sum_{i=1}^N \left\| \phi \frac{\partial v}{\partial y_i} \right\|_{L^2(Y^*)}^2 \right]^{1/2}$$

defines a norm equivalent to the quotient norm on $V/\mathbf{R}$.

Now, we formulate the eigenvalue problems in the spaces V_ϵ . We define

$$(5.11) \quad \begin{cases} a(\phi^\epsilon ; u, v) = \int_{\Omega_\epsilon} \phi^\epsilon \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_i} dx \text{ for } u, v \in V_\epsilon, \\ (\phi^\epsilon ; u, v) = \int_{\Omega_\epsilon} \phi^\epsilon u v dx \forall u, v \in L^2(\phi^\epsilon). \end{cases}$$

By virtue of the properties (4) and (5) of the proposition (5.1), the following problem is well posed:

$$(5.12) \quad \begin{cases} \text{Find } (v_\epsilon, \mu_\epsilon) \in V_\epsilon \times \mathbf{R} \text{ such that} \\ a(\phi^\epsilon ; v_\epsilon, v) = \mu_\epsilon (\phi^\epsilon ; v_\epsilon, v) \text{ for } v \in V_\epsilon, \\ (\phi^\epsilon ; v_\epsilon, v_\epsilon) = 1. \end{cases}$$

Let $\{\mu_\epsilon^i\}_{i=1}^\infty$ be the sequence of eigenvalues and $\{v_\epsilon^i\}_{i=1}^\infty$ the sequence of corresponding vectors satisfying

$$(5.13) \quad \begin{cases} \text{(i) } 0 < \mu_\epsilon^1 \leq \mu_\epsilon^2 \leq \dots \rightarrow \infty, \\ \text{(ii) } \{v_\epsilon^i\}_{i=1}^\infty \text{ form an orthonormal basis in } L^2(\phi^\epsilon). \end{cases}$$

We have once again the characterization of μ_ϵ^1 , analogous to (3.6), in terms of Rayleigh quotient

$$R(\phi^\epsilon ; v) = \frac{a(\phi^\epsilon ; v, v)}{(\phi^\epsilon ; v, v)} \text{ for } v \in V_\epsilon, v \neq 0.$$

6. Estimations on the eigenvalues

The following Lemma establishes one important relation between the eigenvalues λ_ϵ , λ and μ_ϵ .

Lemma (6.1). Let λ be the first eigenvalue of the problem (4.3) with the eigenvector ϕ . Then for all $v \in V_\epsilon$, we have

$$(6.1) \quad \int_{\Omega_\epsilon} \frac{\partial}{\partial x_i} (\phi^\epsilon v) \frac{\partial}{\partial x_i} (\phi^\epsilon v) dx = \epsilon^{-2} \lambda \int_{\Omega_\epsilon} \phi^{\epsilon^2} v^2 dx + \int_{\Omega_\epsilon} \phi^{\epsilon^2} \frac{\partial v}{\partial x_i} \frac{\partial v}{\partial x_i} dx.$$

Proof. Since $D(\Omega_\epsilon)$ is dense in V_ϵ , it suffices to verify (6.1) for all $v \in D(\Omega_\epsilon)$. On the one hand, we have

$$\int_{\Omega_\epsilon} \frac{\partial}{\partial x_i} (\phi^\epsilon v) \frac{\partial}{\partial x_i} (\phi^\epsilon v) dx = \int_{\Omega_\epsilon} \frac{\partial \phi^\epsilon}{\partial x_i} \frac{\partial}{\partial x_i} (\phi^\epsilon v^2) dx + \int_{\Omega_\epsilon} \phi^{\epsilon^2} \frac{\partial v}{\partial x_i} \frac{\partial v}{\partial x_i} dx$$

and on the other hand, by multiplying (4.5) by $\phi^\epsilon v^2$ we obtain

$$\int_{\Omega_\epsilon} \frac{\partial \phi^\epsilon}{\partial x_i} \frac{\partial}{\partial x_i} (\phi^\epsilon v^2) dx = \epsilon^{-2} \lambda \int_{\Omega_\epsilon} \phi^{\epsilon^2} v^2 dx,$$

and so we deduce (6.1) without difficulty.

Corollary (6.1)

Let $\{\lambda_\epsilon^l\}_{l=1}^\infty$, $\{\mu_\epsilon^l\}_{l=1}^\infty$ be the sequences of eigenvalues of the problems (3.2) and (5.12) respectively. Let λ be the first eigenvalue of the problem (4.3). Then we have

$$(6.2) \quad \lambda_\epsilon^l = \epsilon^{-2} \lambda + \mu_\epsilon^l \text{ for } l \geq 1.$$

Proof. In fact, we obtain, from (6.1)

$$R_\epsilon(\phi^\epsilon v) = \epsilon^{-2} \lambda + R(\phi^\epsilon; v) \text{ for } v \in V_\epsilon.$$

Now we use the minimax principles for eigenvalues and the isomorphism of the proposition (5.1) (iii) to get the relation (6.2).

Since $\mu_\epsilon^l > 0$ for all $l \geq 1$ and $\epsilon > 0$, we see that the sequence $\{\lambda_\epsilon^l - \epsilon^{-2} \lambda\}_{\epsilon > 0}$ is bounded below by zero. The following Proposition shows that it is bounded above.

Proposition (6.1). Let $\{\mu_\epsilon^l\}_{l=1}^\infty$ be the sequence of eigenvalues of the problem (5.12). Then for $l \geq 1$, $\{\mu_\epsilon^l\}_{\epsilon > 0}$ is bounded independently of ϵ .

Proof. We use the following characterizations of minimax principle :

$$(6.4) \quad \mu_\epsilon^l = \min_{v \in S_l} \{ \max_{v \in S_l} R(\phi^\epsilon; v) ; S_l \subset V_\epsilon, \dim S_l = l \}.$$

We take S_l to be the vector space spanned by $w_1, w_2, \dots, w_l$, the first l -eigenvectors of the following Dirichlet problem :

$$(6.5) \quad \begin{cases} \text{Find } (w, v) \in H_0^1(\Omega) \times \mathbf{R} \text{ such that} \\ -\Delta w = vw \text{ in } \Omega. \end{cases}$$

It is not difficult to see that we have

$$(6.6) \quad \dim(S_l/\Omega_\epsilon) = l \text{ for } \epsilon > 0.$$

$[S_i/\Omega_\epsilon]$ denotes the restriction to Ω_ϵ of functions belonging to S_i . So, one can take S_i/Ω_ϵ in (6.4). (This is alright, since no boundary condition is required on S_ϵ for V_ϵ .) We obtain

$$(6.7) \quad \mu_\epsilon^1 \leq \max_{v \in S_1} R(\phi^\epsilon; v).$$

We claim that the right hand side of the inequality (6.7) is bounded above by a constant independent of ϵ . In fact, on the contrary we would have, for a sequence $\epsilon_n \rightarrow 0$, a sequence $\{v_n\} \subset S_1$ such that

$$(6.8) \quad \int_{\Omega_n} \phi_n^2 \nabla v_n \nabla v_n > n \int_{\Omega_n} \phi_n^2 v_n^2 \text{ for } n \geq 1,$$

$$(6.9) \quad \int_{\Omega} v_n^2 = 1 \text{ for } n \geq 1.$$

Here we have set $\Omega_{\epsilon_n} = \Omega_n$ and $\phi^{\epsilon_n} = \phi_n$. Since S_1 is of finite dimension, we have (for a subsequence)

$$v_n \rightarrow v \text{ in } H_0^1(\Omega) \text{ strong.}$$

Now, one can pass to the limit in (6.8) and obtain

$$m_1(\phi^2) \int_{\Omega} v^2 = 0.$$

But (6.9) implies that

$$\int_{\Omega} v^2 = 1.$$

This contradiction proves the Proposition.

7. Asymptotic development

The aim of this section is to find "the homogenized operator" for the problem (3.2) by the method of asymptotic development introduced in Bensoussan *et al* [3] and Lions [13].

We introduce one "fast" variable :

$$(7.1) \quad y = x/\epsilon.$$

Then, the differential operator $\partial/\partial x_j$ applied to a function $\phi(x, y)$ becomes

$$(7.2) \quad \frac{\partial}{\partial x_j} + \epsilon^{-1} \frac{\partial}{\partial y_j}.$$

So, the laplacian operator is transformed into

$$(7.3) \quad \epsilon^{-2} \Delta_y + 2\epsilon^{-1} \Delta_{yy} + \Delta_x$$

where

$$(7.4) \quad \Delta_y = \frac{\partial^2}{\partial y_i \partial y_i}, \quad \Delta_{yy} = \frac{\partial^2}{\partial x_i \partial y_i}, \quad \Delta_x = \frac{\partial^2}{\partial x_i \partial x_i}.$$

Taking into account Proposition (6.1) and Corollary (6.1), we propose the following Ansatz for the problem (3.1) :

$$(7.5) \quad u_\epsilon(x) = u_0(x, y) + \epsilon u_1(x, y) + \dots, \quad y = x/\epsilon,$$

$$(7.6) \quad \lambda_\epsilon = \epsilon^{-2} \lambda_{-2} + \epsilon^{-1} \lambda_{-1} + \lambda_0 + \dots$$

We impose the following restrictions on the functions u_j which are defined for $x \in \Omega$ and $y \in Y^*$:

$$(7.7) \quad \begin{cases} u_j(x, y) = 0 & \text{if } y \in S, \\ u_j(x, \cdot) & \text{is } Y\text{-periodic in } y. \end{cases}$$

We substitute the expressions (7.5) and (7.6) in equation (3.1) and we identify the powers of ϵ . We obtain

$$(7.8) \quad \begin{cases} u_0(x, y) = \phi(y) \tilde{u}_0(x), \\ u_1(x, y) = \phi^{(j)}(y) \frac{\partial \tilde{u}_0}{\partial x_j}(x) + \phi(y) \tilde{u}_1(x), \end{cases}$$

$$(7.9) \quad \lambda_{-2} = \lambda,$$

$$(7.10) \quad \lambda_{-1} = 0,$$

$$(7.11) \quad A \tilde{u}_0 = \lambda_0 \tilde{u}_0,$$

where

(1) λ is the first eigenvalue and ϕ the corresponding eigenvector of the problem (4.3).

(2) The functions $\phi^{(j)}$ ($j = 1, 2, \dots, N$) are defined by

$$(7.12) \quad a(\phi^{(j)}, v) - \lambda(\phi^{(j)}, v)_{Y^*} = 2 \left(\frac{\partial \phi}{\partial y_j}, v \right)_{Y^*} \quad \text{for } v \in W_0, \phi^{(j)} \in W_0.$$

(3) The operator A (called homogenized operator) is defined by

$$(7.13) \quad A \equiv -q_{ij} \frac{\partial^2}{\partial x_i \partial x_j},$$

the "homogenized coefficients" being defined by

$$(7.14) \quad q_{ij} = \delta_{ij} + 2 \int_{Y^*} \frac{\partial \phi^{(j)}}{\partial y_i} \phi \, dy \quad \text{for } i, j = 1, 2, \dots, N.$$

Remark (7.1)

The relations (7.9) and (7.10) are in accordance with the results of Proposition (6.1) and Corollary (6.1).

Remark (7.2) Since

$$\left(\frac{\partial \phi}{\partial y_i}, \phi \right)_{Y^*} = 0 \quad \text{for } i = 1, 2, \dots, N,$$

equations (7.12) can be solved for $\phi^{(j)}$ by Fredholm alternative,

The "homogenized problem" is an eigenvalue problem for the operator A :

$$(7.15) \quad \begin{cases} \text{Find } (u, \mu) \in H_0^1(\Omega) \times \mathbf{R} \text{ such that} \\ Au = \mu u \text{ in } \Omega, \\ u \neq 0. \end{cases}$$

The preceding formal calculations show that the sequence $\{\lambda_\epsilon - \epsilon^{-2} \lambda\}_{\epsilon > 0}$ converges to an eigenvalue of the problem (7.15). We prove this result later by the method of energy.

8. Ellipticity of the homogenized operator

The idea of proving the ellipticity of the operator consists of identifying the coefficients q_{ij} with the homogenized coefficients associated with the problem (5.12). So, we apply the asymptotic development method to the problem (5.12). First, we write the problem (5.12) in operator form : The solution $(v_\epsilon, \mu_\epsilon)$ is characterized by

$$(8.1) \quad \begin{cases} (v_\epsilon, \mu_\epsilon) \in V_\epsilon \times \mathbf{R}, \\ -\frac{\partial}{\partial x_i} \left(\phi^{\epsilon^2} \frac{\partial v_\epsilon}{\partial x_i} \right) = \mu_\epsilon \phi^{\epsilon^2} v_\epsilon \text{ in } \Omega_\epsilon, \\ \int_{\Omega_\epsilon} \phi^{\epsilon^2} v_\epsilon^2 = 1. \end{cases}$$

We develop v_ϵ and μ_ϵ in the following form :

$$(8.2) \quad v_\epsilon(x) = v_0(x, y) + \epsilon v_1(x, y) + \dots, \quad y = x/\epsilon,$$

$$(8.3) \quad \mu_\epsilon = \mu_0 + \epsilon \mu_1 + \dots,$$

where v_j is defined on $\Omega \times Y^*$ and it is Y -periodic in y .

We put these expressions in (8.1) and identify the powers of ϵ . We get the following results :

$$(8.4) \quad v_0(x, y) = \tilde{v}_0(x),$$

$$(8.5) \quad v_1(x, y) = \psi^{(j)}(y) \frac{\partial \tilde{v}_0}{\partial x_j}(x) + \tilde{v}_1(x),$$

and the necessary and sufficient condition so that we solve for v_2 is

$$(8.6) \quad B\tilde{v}_0 = \mu_0 \tilde{v}_0 \text{ in } \Omega,$$

where the operator B is defined by

$$(8.7) \quad B \equiv -p_{ij} \frac{\partial^2}{\partial x_i \partial x_j},$$

with

$$(8.8) \quad p_{ij} = \delta_{ij} + \int_{Y^*} \phi^2 \frac{\partial \psi^{(i)}}{\partial y_i} dy \text{ for } i, j = 1, 2, \dots, N.$$

The functions $\psi^{(j)}$ ($j = 1, 2, \dots, N$) are defined (upto an additive constant) to be solutions of the following variational problem :

$$(8.9) \quad \begin{cases} a(\phi; \psi^{(j)}, v) = -a(\phi; y_j, v) \forall v \in V, \\ \psi^{(j)} \in V, \end{cases}$$

the bilinear form $a(\phi; \cdot, \cdot)$ being defined by

$$(8.10) \quad a(\phi; u, v) = \int_{Y^*} \phi^2 \frac{\partial u}{\partial y_i} \frac{\partial v}{\partial y_i} dy \quad \forall u, v \in V.$$

It is not difficult to see that the operator B is elliptic in the sense that there exists a constant $\alpha > 0$ such that

$$(8.11) \quad p_{ij} \xi_i \xi_j \geq \alpha \xi_i \xi_i \text{ for } \xi = (\xi_i) \in \mathbb{R}^N.$$

In fact, the coefficients p_{ij} can be expressed by the following formula :

$$(8.12) \quad p_{ij} = a(\phi; \psi^{(j)} + y_j, \psi^{(i)} + y_i) \text{ for } i, j = 1, 2, \dots, N,$$

taking into account the following relation

$$a(\phi; \psi^{(j)} + y_j, \psi^{(i)}) = 0 \text{ for } i, j = 1, 2, \dots, N.$$

Now, the inequality (8.11) is a simple consequence of (8.12).

Theorem (8.1)

Let (q_{ij}) and (p_{ij}) be defined by (7.14) and (8.8) respectively. Then we have

$$q_{ij} = p_{ij} \text{ for } i, j = 1, 2, \dots, N.$$

In particular, the operator A is symmetric and elliptic.

Proof. We prove that $\phi\psi^{(j)}$ is a solution to the problem (7.12). For that we use the isomorphism between the spaces V and W_0 given by the Proposition (5.2) (iii). Firstly, we have $\phi\psi^{(j)} \in W_0$.

By virtue of the Proposition (5.2) (i), it suffices to verify that

$$(8.15) \quad \int_{Y^*} \frac{\partial}{\partial y_i} (\phi\psi^{(j)}) \frac{\partial}{\partial y_i} (\phi v) dy = \lambda \int_{Y^*} \phi^2 \psi^{(j)} v dy + 2 \int_{Y^*} \frac{\partial \phi}{\partial y_j} \phi v dy,$$

for all $v \in c^\infty(\bar{Y}^*) \cap V$ which vanish in a neighbourhood of S . The relation (8.15) is easily proved by using the definitions of ϕ and $\psi^{(j)}$ and the following identity :

$$(8.16) \quad 2 \int_{Y^*} \phi \frac{\partial \phi}{\partial y_i} v dy = - \int_{Y^*} \phi^2 \frac{\partial v}{\partial y_i} dy \quad \forall v \in V.$$

Now we take $\phi^{(j)} = \phi\psi^{(j)}$ in the formula (7.14) defining q_{ij} . We obtain

$$q_{ij} = \delta_{ij} + 2 \int_{Y^*} \frac{\partial}{\partial y_i} (\phi\psi^{(j)}) \phi dy,$$

$$= \delta_{ij} + \int_{Y^*} \phi^2 \frac{\partial \psi^{(i)}}{\partial y_i} dy$$

$$= p_{ij},$$

and so the Theorem is proved.

9. Homogenization theorems

In this section using energy method we show how one passes to the limit as $\epsilon \rightarrow 0$ in the problem (5.12) which determines the correctors μ_ϵ . Before that, we need some more notations. Let us denote by T' any hole satisfying the hypothesis (H5). We denote by $P_{\epsilon'}^{T'}$, the prolongation operator constructed from the hole T' satisfying the following condition :

$$(9.1) \quad \left\{ \begin{array}{l} \text{For all } v \text{ in } H^1(\Omega_{T'_\epsilon}) \text{ with } v = 0 \text{ on } \partial\Omega, \text{ we have} \\ P_{\epsilon'}^{T'} v \in H_0^1(\Omega) \text{ and there exists a constant } c > 0 \\ \text{independent of } \epsilon \text{ (but depending on the hole } T') \text{ such that} \\ |P_{\epsilon'}^{T'} v|_{1, \Omega} \leq c |v|_{1, \Omega_{T'_\epsilon}}, \\ \text{where } |v|_{1, \Omega} = \int_{\Omega} \nabla v \cdot \nabla v \, dx. \end{array} \right.$$

The existence of the operator $P_{\epsilon'}^{T'}$ is proved in Cioranescu and Saint Jean Paulin [8] and Cioranescu [7].

Lemma (9.1). Suppose for each $\epsilon > 0$, we are given w_ϵ in V_ϵ such that

$$(9.2) \quad \|w_\epsilon\|_\epsilon \leq c, \text{ independent of } \epsilon.$$

Then there exists a subsequence of ϵ (again denoted by ϵ) and a function w_0 in $H_0^1(\Omega)$ such that

$$(9.3) \quad P_{\epsilon'}^{T'} w_\epsilon \rightarrow w_0 \text{ in } H_0^1(\Omega) \text{ weak,}$$

for all holes T' verifying (H5).

Proof. It follows from the hypothesis (9.2) and from the fact that $\phi > 0$ in Y^* that there exists a constant $c_{T'}$ such that

$$(9.4) \quad |w_\epsilon|_{1, \Omega_{T'_\epsilon}} \leq c_{T'}.$$

Thanks to the inequality (9.1), we see that the sequence $\{P_{\epsilon'}^{T'} w_\epsilon\}$ remains bounded in $H_0^1(\Omega)$. So, we can extract a subsequence of ϵ and a function $w_0^{T'}$ in $H_0^1(\Omega)$ both depending on T' such that

$$(9.5) \quad P_{\epsilon'}^{T'} w_\epsilon \rightarrow w_0^{T'} \text{ in } H_0^1(\Omega) \text{ weak,}$$

for the extracted subsequence.

Now let us consider another hole satisfying (H5). By similar arguments as before, we obtain a subsequence of the subsequence already picked up and a function $w_0^{T''}$ in $H_0^1(\Omega)$ such that

$$(9.6) \quad P_{\epsilon}^{T''} w_{\epsilon} \rightarrow w_0^{T''} \text{ in } H_0^1(\Omega) \text{ weak.}$$

But we have the relation

$$(9.7) \quad \chi_{Y-(T' \cup T'')} (x/\epsilon) P_{\epsilon}^{T''} w_{\epsilon} = \chi_{Y-(T' \cup T'')} (x/\epsilon) P_{\epsilon}^{T''} w_{\epsilon} \text{ in } \Omega$$

where $\chi_{Y-(T' \cup T'')} (y)$ is the characteristic function of $Y - (T' \cup T'')$. One can pass to the limit in (9.7) and we obtain

$$m_Y (\chi_{Y-(T' \cup T'')} w_0^{T'}) = m_Y (\chi_{Y-(T' \cup T'')} w_0^{T''}),$$

and so $w_0^{T'} = w_0^{T''}$ in Ω . By the uniqueness of the limit, we see that the proof is complete.

Lemma (9.2). Let $\{w_{\epsilon}\}_{\epsilon} > 0$ be given as in the preceding Lemma. Then for the subsequence of ϵ and for the function w_0 in $H_0^1(\Omega)$ given by the Lemma (9.1) we have

$$(9.8) \quad P_{\epsilon}^{T'} w_{\epsilon} \rightarrow w_0 \text{ in } H_0^1(\Omega) \text{ weak for all holes } T',$$

$$(9.9) \quad \int_{\Omega_{\epsilon}} \phi (x/\epsilon) f (x/\epsilon) w_{\epsilon} dx \rightarrow \int_{\Omega} m_Y (\phi f) w_0 dx,$$

where f is given in $L^2(Y)$.

Proof. Let T' be a hole satisfying (H5). We write

$$\begin{aligned} (9.10) \quad & \int_{\Omega_{\epsilon}} \phi^{\epsilon} f^{\epsilon} w_{\epsilon} dx - \int_{\Omega} m_Y (\phi f) w_0 dx \\ &= \left[\int_{\Omega_{\epsilon} - \Omega_{T'_{\epsilon}}} \phi^{\epsilon} f^{\epsilon} w_{\epsilon} dx - \int_{\Omega} m_{Y-T'} (\phi f) w_0 \right] + \left[\int_{\Omega_{T'_{\epsilon}} - \Omega_{T'_{\epsilon}'}} \phi^{\epsilon} f^{\epsilon} w_{\epsilon} dx \right. \\ & \quad \left. - \int_{\Omega} m_{T'-T} (\phi f) w_0 dx \right], \end{aligned}$$

where we have set

$$(9.11) \quad m_U (g) = \frac{1}{|Y|} \int_U g dy \text{ for subsets } U \text{ of } Y.$$

We have the following estimate :

$$\left| \int_{\Omega_{T'_{\epsilon}} - \Omega_{T'_{\epsilon}'}} \phi^{\epsilon} f^{\epsilon} w_{\epsilon} dx \right| \leq \| \phi \|_{L^{\infty}(T'-T)} \| f^{\epsilon} \|_{L^2(\Omega_{\epsilon})} \| w_{\epsilon} \|_{L^2(\Omega_{\epsilon})}.$$

Using now Proposition (5.1) (2), we deduce

$$(9.12) \quad \left| \int_{\Omega_{T'_{\epsilon}} - \Omega_{T'_{\epsilon}'}} \phi^{\epsilon} f^{\epsilon} w_{\epsilon} dx \right| \leq c \| \phi \|_{L^{\infty}(T'-T)},$$

where c is a constant independent of T' and ϵ .

We also have

$$(9.13) \quad \left| \int_{\Omega} m_{T'-T} (\phi f) w_0 dx \right| \leq c \| \phi \|_{L^{\infty}(T'-T)}.$$

Since $\phi = 0$ on S , we can choose T' satisfying (H5) and $\|\phi\|_{L^\infty(T'-T)}$ sufficiently small. For this hole T' , we have the convergence

$$(9.14) \quad \int_{\Omega_{T'_\epsilon}} \phi^\epsilon f^\epsilon P_\epsilon^{T'} w_\epsilon dx \rightarrow \int_{\Omega} m_{Y-T'}(\phi f) w_0 dx.$$

Combining all these results, we obtain (9.9).

Using similar arguments, we can prove the following :

Lemma (9.3). Let $\{w_\epsilon\}$ and $\{v_\epsilon\}$ be two sequences such that

$$(9.15) \quad \left. \begin{array}{l} \|w_\epsilon\|_\epsilon \leq c \\ \|v_\epsilon\|_\epsilon \leq c \end{array} \right\} \text{ independent of } \epsilon.$$

Then there exist a subsequence of ϵ (again denoted by ϵ) and functions w_0, v_0 in $H_0^1(\Omega)$ such that we have the following convergence for this subsequence of ϵ :

$$(9.16) \quad \left\{ \begin{array}{l} P_\epsilon^{T'} w_\epsilon \rightarrow w_0 \text{ in } H_0^1(\Omega) \text{ weak,} \\ P_\epsilon^{T'} v_\epsilon \rightarrow v_0 \text{ in } H_0^1(\Omega) \text{ weak,} \end{array} \right.$$

for all holes T' satisfying (H5) and

$$(9.17) \quad \int_{\Omega_\epsilon} \phi^\epsilon f^\epsilon w_\epsilon v_\epsilon dx \rightarrow \int_{\Omega} m_Y(\phi f) w_0 v_0 dx,$$

for all f in $L^\infty(Y)$.

Now we have all the tools to prove the following homogenization theorem:

Theorem (9.1). We suppose that (H5) is satisfied. Let $(v_\epsilon, \mu_\epsilon)$ be a solution of the problem (5.12). We assume

$$(9.18) \quad \mu_\epsilon \rightarrow \mu \text{ as } \epsilon \rightarrow 0.$$

Then μ is an eigenvalue of the "homogenized problem" (7.15). Furthermore there is an eigenvector v_0 in $H_0^1(\Omega)$ corresponding to μ and a subsequence of ϵ (again denoted by ϵ) such that

$$(9.19) \quad P_\epsilon^{T'} v_\epsilon \rightarrow v_0 \text{ in } H_0^1(\Omega) \text{ weak for all holes } T'$$

satisfying (H5), $P_\epsilon^{T'}$ being the prolongation operator verifying (9.1).

Proof.

Step 1. Taking $v = v_\epsilon$ in (5.12), we obtain

$$(9.20) \quad \|v_\epsilon\|_\epsilon \leq c, \text{ independent of } \epsilon.$$

Set

$$(9.21) \quad \xi_i^\epsilon = \phi^\epsilon \frac{\partial v_\epsilon}{\partial x_i} \text{ in } \Omega_\epsilon, i = 1, 2, \dots, N.$$

It follows that

$$(9.22) \quad \|\xi_i^\epsilon\|_{L^2(\Omega_\epsilon)} \leq c, i = 1, 2, \dots, N,$$

and we have

$$(9.23) \quad -\operatorname{div}(\phi^\epsilon \xi^\epsilon) = \mu_\epsilon \phi^{\epsilon^2} v_\epsilon \text{ in } \Omega_\epsilon.$$

So, $\phi^\epsilon \xi^\epsilon \cdot v_\epsilon$ admits a trace on $\partial\Omega_\epsilon$ where v_ϵ is the exterior normal to $\partial\Omega_\epsilon$. It is not difficult to see that

$$(9.24) \quad \phi^\epsilon \xi^\epsilon \cdot v_\epsilon = 0 \text{ on } S_\epsilon.$$

So, equation (9.23) is valid in Ω if we extend ϕ^ϵ by zero in the holes.

$$(9.25) \quad -\operatorname{div}(\phi^\epsilon \xi^\epsilon) = \mu_\epsilon \phi^{\epsilon^2} v_\epsilon \text{ in } \Omega.$$

In (9.25), we take some arbitrary extensions for ξ^ϵ and v_ϵ . Using Lemma (9.1), we can extract a subsequence of ϵ (again denoted by ϵ) such that

$$(9.26) \quad \phi^\epsilon \xi_i^\epsilon \rightarrow \xi_i \text{ in } L^2(\Omega) \text{ weak, } i = 1, 2, \dots, N.$$

$$(9.27) \quad P_\epsilon^{T'} v_\epsilon \rightarrow v_0 \text{ in } H_0^1(\Omega) \text{ weak, for all holes}$$

T' satisfying (H5).

Step 2. Using the technique of Lemma (9.2), we can pass to the limit in (9.25) to get

$$(9.28) \quad -\operatorname{div} \xi = \mu m_Y(\phi^2) v_0 \text{ in } \Omega.$$

Step 3. We introduce w as solution to the following variational problem :

$$(9.29) \quad \begin{cases} a(\phi; w, v) = 0 \text{ for } v \in V, \\ w - \Pi \in V, \end{cases}$$

where $\Pi(y)$ is a homogeneous polynomial of degree 1. Set

$$(9.30) \quad \eta = \phi \nabla w \text{ in } Y^*.$$

Then η satisfies

$$(9.31) \quad -\operatorname{div}(\phi \eta) = 0 \text{ in } Y^*.$$

Also we have $\phi \eta \cdot v = 0$ on S where v is the outer normal to S . So, we obtain

$$(9.32) \quad -\operatorname{div}(\phi^\epsilon \eta^\epsilon) = 0 \text{ in } \Omega.$$

We set

$$\psi(y) = w(y) - \Pi(y) \text{ for } y \in Y^*.$$

Then $\psi \in V$ and we take some extension of ψ in the hole and we extend ψ periodically throughout $\mathbf{R}^N$. We define

$$(9.33) \quad \begin{aligned} w^\epsilon(x) &= \epsilon w(x/\epsilon) \\ &= \Pi(x) + \epsilon \psi(x/\epsilon) \text{ in } \Omega. \end{aligned}$$

Since $\psi \in L^2(Y)$, $\psi(x/\epsilon)$ remains bounded in $L^2(\Omega)$ and so

$$(9.34) \quad w^\epsilon \rightarrow \Pi \text{ in } L^2(\Omega) \text{ strong.}$$

Step 4. Let $g \in D(\Omega)$.

We take $v = gw^\epsilon$ in (5.12) and multiply (9.32) by gv_ϵ and subtract, we get

$$(9.35) \quad \int_{\Omega_\epsilon} \phi^\epsilon \xi_i^\epsilon w^\epsilon \frac{\partial g}{\partial x_i} - \int_{\Omega_\epsilon} \phi^\epsilon \eta_i^\epsilon v_\epsilon \frac{\partial g}{\partial x_i} = \mu_\epsilon \int_{\Omega_\epsilon} \phi^{\epsilon^2} v_\epsilon g w^\epsilon dx.$$

Using the technique of Lemma (9.2), we pass to the limit in (9.35) to get

$$(9.36) \quad \int_{\Omega} \xi_i \Pi \frac{\partial g}{\partial x_i} - \int_{\Omega} m_Y(\phi \eta_i) v_0 \frac{\partial g}{\partial x_i} = \mu m_Y(\phi^2) \int_{\Omega} v_0 g \Pi dx.$$

It follows from (9.36) and (9.28) that

$$(9.37) \quad \xi_i \frac{\partial \Pi}{\partial x_i} = m_Y \left(\phi^2 \frac{\partial w}{\partial y_i} \right) \frac{\partial v_0}{\partial x_i} \text{ in } \Omega.$$

Step 5. Now we take $\Pi = y_j$ ($j = 1, 2, \dots, N$). The corresponding test function $w = \psi^{(j)} + y_j$ where $\psi^{(j)}$ is given by (8.9). So, we get

$$(9.38) \quad \xi_j = m_Y(\phi^2) q_{ij} \frac{\partial v_0}{\partial x_i} \text{ for } j = 1, 2, \dots, N.$$

This when combined with (9.28) shows that

$$(9.39) \quad Av_0 = \mu v_0 \text{ in } \Omega.$$

Step 6. To complete the proof, we have to show that

$$(9.40) \quad v_0 \neq 0.$$

We have

$$(9.41) \quad \int_{\Omega_e} \phi^{\epsilon^2} v_e^2 = 1,$$

and at the limit we obtain

$$m_Y(\phi^2) \int_{\Omega} v_0^2 = 1,$$

and so (9.40) follows.

Theorem (9.2). We assume (H5). Let λ be the first eigenvalue of the problem (4.3) with eigenvector ϕ . Let $\{v_e^l, \mu_e^l\}_{l=1}^{\infty}$ be the spectrum of the problem (5.12). Let $\{\lambda_e^l\}_{l=1}^{\infty}$ be the sequence of eigenvalue of the problem (3.1). Then

- (i) $\lambda_e^l - \epsilon^{-2} \lambda = \mu_e^l$ for $l \geq 1$.
- (ii) $\mu_e^l \rightarrow l$ th eigenvalue μ^l of the problem (7.15).
- (iii) There exist a subsequence of ϵ (again denoted by ϵ) and eigenvectors $\{v^l\}$ of the problem (7.15) corresponding to $\{\mu^l\}$ such that

$$(9.42) \quad P_{\epsilon}^{T'} v_e \rightarrow v^l \text{ in } H_0^1(\Omega) \text{ weak for all holes } T' \text{ satisfying (H5).}$$

$$(9.43) \quad \{v^l\} \text{ form an orthogonal base in } L^2(\Omega).$$

- (iv) If μ^l is a simple eigenvalue of (7.15), then given any eigenvector v_0^l corresponding to μ^l satisfying (9.42), we can choose an eigenvector $v_{e,0}^l$ of the problem (5.12) corresponding to μ_e^l such that

$$(9.44) \quad P_{\epsilon}^{T'} v_{e,0}^l \rightarrow v_0^l \text{ in } H_0^1(\Omega) \text{ weak for the whole sequence of } \epsilon.$$

Proof. By Proposition (6.1), we can extract a subsequence of ϵ such that

$$(9.45) \quad \mu_\epsilon^l \rightarrow \mu^l \text{ as } \epsilon \rightarrow 0.$$

According to Theorem (9.1), μ^l is an eigenvalue of the problem (7.15) with eigenvector v^l satisfying (9.42). To prove μ^l is the l th eigenvalue, it suffices to verify that

- (1) $\{v^l\}$ is an orthogonal base in $L^2(\Omega)$.
- (2) There is no other eigenvalue except $\{\mu^l\}$ for the problem (7.15) : $(u, \mu) \in H_0^1(\Omega) \times \mathbb{R}$ such that $Au = \mu u$.

We remark, first, that the eigenvectors $\{v^l\}$ are orthogonal in $L^2(\Omega)$. In fact, the passage to the limit in the relation

$$(9.46) \quad \int_{\Omega_\epsilon} \phi^{\epsilon^2} v_\epsilon^l v_\epsilon^m = \delta_{lm},$$

will give

$$(9.47) \quad m_Y(\phi^2) \int_{\Omega} v^l v^m = \delta_{lm}.$$

We shall prove (2). Let μ be an eigenvalue of the problem (7.15) which is different from μ^l with eigenvector w satisfying

$$(9.48) \quad \begin{cases} Aw = \mu w \text{ in } \Omega, \\ \mu \neq \mu^l \text{ for any } l, \\ \int_{\Omega} w^2 = 1/(m_Y(\phi^2)), \\ \int_{\Omega} w v^l = 0 \text{ for each } l \geq 1. \end{cases}$$

We can choose l such that

$$(9.49) \quad \mu < \mu^{l+1}.$$

Now, we define w_ϵ as follows :

$$(9.50) \quad \begin{cases} w_\epsilon \in V_\epsilon \\ a(\phi^\epsilon; w_\epsilon, v) = \mu(\phi^\epsilon; w, v) \quad \forall v \in V_\epsilon. \end{cases}$$

The proof of Theorem (9.1) shows that

$$(9.51) \quad P_\epsilon^{T'} w_\epsilon \rightarrow w \text{ in } H_0^1(\Omega) \text{ weak.}$$

Set

$$(9.52) \quad \hat{w}_\epsilon = w_\epsilon - \sum_{i=1}^l (\phi^\epsilon; w_\epsilon, v_\epsilon^i) v_\epsilon^i.$$

We see easily that

$$(9.53) \quad (\phi^\epsilon; \hat{w}_\epsilon, v_\epsilon^i) = 0 \text{ for } i = 1, 2, \dots, l,$$

and as a consequence

$$(9.54) \quad a(\phi^\epsilon; \hat{w}_\epsilon, \hat{w}_\epsilon) \geq \mu_\epsilon^{l+1}(\phi^\epsilon; \hat{w}_\epsilon, \hat{w}_\epsilon).$$

We can prove

$$(9.55) \quad \begin{cases} a(\phi^\epsilon; \hat{w}_\epsilon, \hat{w}_\epsilon) \rightarrow \mu, \\ (\phi^\epsilon; \hat{w}_\epsilon, \hat{w}_\epsilon) \rightarrow m_Y(\phi^2) \int_{\Omega} w^2 = 1. \end{cases}$$

Hence (9.54) implies that

$$(9.56) \quad \mu \geq \mu^{l+1},$$

which is in contradiction with (9.49). This completes the proof of (2). In the same way, one can prove $v \in L^2(\Omega)$, $\int_{\Omega} v v^l = 0$ for all l implies $v = 0$. This proves that v^l is the l th eigenvalue of the problem (7.15).

To prove (iv). We note, first, that μ_ϵ^l is simple for ϵ sufficiently small. In fact, since we have $\mu_\epsilon^l \rightarrow \mu^l$ for all l , the multiplicity of μ_ϵ^l is less than or equal to that of μ^l . So, there are only two vectors $v_{\epsilon,0}^l$ and $-v_{\epsilon,0}^l$ which satisfy (5.12). We choose one which satisfies

$$(9.57) \quad (\phi^\epsilon; v_{\epsilon,0}^l, v_0^l) > 0,$$

and this sequence must satisfy (9.44).

Part B: Stekloff eigenvalue problem

10. Problem to be treated

With the notations introduced in § 2, we consider the following eigenvalue problem:

$$(10.1) \quad \begin{cases} \text{Find } (u_\epsilon, \lambda_\epsilon) \in \mathcal{W}_\epsilon \times \mathbb{R} \text{ such that} \\ -\Delta u_\epsilon = 0 \text{ in } \Omega_\epsilon, \\ \frac{\partial u_\epsilon}{\partial \nu_\epsilon} = \lambda_\epsilon u_\epsilon \text{ on } S_\epsilon, \\ u_\epsilon \neq 0, \end{cases}$$

where $\partial/\partial \nu_\epsilon$ is the exterior normal derivative to S_ϵ and

$$(10.2) \quad \mathcal{W}_\epsilon = \{v \in H^1(\Omega_\epsilon); v = 0 \text{ on } \partial\Omega\}.$$

Problems of the type (10.1) have been studied by Bergmann and Schiffer [4] and Bramble and Osborn [6]. This problem can be put in the variational form as follows: Consider the following problem for $g \in L^2(S_\epsilon)$:

$$\begin{cases} a^\epsilon(w_\epsilon, v) = \int_{S_\epsilon} g v \text{ for } v \in \mathcal{W}_\epsilon, \\ w_\epsilon \in \mathcal{W}_\epsilon. \end{cases}$$

Set $G_\epsilon g = w_{\epsilon|S_\epsilon}$. Then G_ϵ is self-adjoint, compact operator in $L^2(S_\epsilon)$. We can write the eigenvalue problem

$$G_\epsilon u_\epsilon = \mu_\epsilon u_\epsilon, \mu_\epsilon = \frac{1}{\lambda_\epsilon}$$

as follows :

$$(10.2) \quad \begin{cases} a^\epsilon(u_\epsilon, v) = \lambda_\epsilon \int_{S_\epsilon} u_\epsilon v \text{ for } v \text{ in } W_\epsilon, \\ u_\epsilon \in W_\epsilon, \\ u_\epsilon \neq 0. \end{cases}$$

We denote by $\{\lambda_{\epsilon l}^l\}_{l=1}^\infty$ the sequence of eigenvalues and by $\{u_\epsilon^l\}$ corresponding eigenvectors of this problem. We remark that $\{u_\epsilon^l\}$ form an orthogonal basis in $L^2(S_\epsilon)$. Our aim, in this part, is to study the asymptotic behaviour of λ_ϵ^l and $\{u_\epsilon^l\}$ as $\epsilon \rightarrow 0$. We prove that $\{\epsilon^{-1} \lambda_\epsilon^l\} \rightarrow l$ th eigenvalue of the "homogenized problem". A very useful tool in this study will be a test function which will be defined in §11.

11. Test function

We define ψ as the solution of the following variational problem :

$$(11.1) \quad \begin{cases} a(\psi, v) = -c_0 \int_{Y^*} v + \int_S v d\gamma \text{ for } v \in W, \\ \psi \in W, \end{cases}$$

where

$$(11.2) \quad W = \{v \in H^1(Y^*); v \text{ is } Y \text{ periodic}\}, \text{ and}$$

$$(11.3) \quad c_0 = \frac{|S|}{|Y^*|}.$$

We define

$$(11.4) \quad \psi^\epsilon(x) = \psi(x/\epsilon) \text{ for } x \in \Omega_\epsilon,$$

and this will satisfy

$$(11.5) \quad \begin{cases} -\Delta \psi^\epsilon = c_0 \epsilon^{-2} \text{ in } \Omega_\epsilon, \\ \frac{\partial \psi^\epsilon}{\partial \nu_\epsilon} = \epsilon^{-1} \text{ on } S_\epsilon, \end{cases}$$

and so, by Green's formula we obtain

$$(11.6) \quad \int_{S_\epsilon} g d\gamma = \int_{\Omega_\epsilon} \frac{\partial \psi}{\partial y_i} (x/\epsilon) \frac{\partial g}{\partial x_i} (x) + c_0 \epsilon^{-1} \int_{\Omega_\epsilon} g \text{ for } g \in W_\epsilon.$$

12. Estimates on the eigenvalues

Proposition (12.1). Let $\{\lambda_\epsilon^l\}$ be the sequence of eigenvalues of the problem (10.1). Then there exists a constant $c_l > 0$ independent of ϵ , such that

$$(12.1) \quad 0 < \lambda_\epsilon^l \leq c_l \epsilon \text{ for } \epsilon > 0.$$

Proof. We make use of the mini-max principle for the eigenvalues :

$$(12.2) \quad \lambda_\epsilon^l = \text{Min} \left\{ \text{Max}_{v \in S_l} \frac{\int_{\Omega_\epsilon} \nabla v \cdot \nabla v}{\int_{S_\epsilon} v^2} ; \quad S_l \subset \mathcal{W}_\epsilon^l \right\}.$$

As in Proposition (6.1), we take S_l to be the space spanned by the first l -eigenvectors of the problem $-\Delta w = \nu w$ in Ω , $w|_{\partial\Omega} = 0$. We get

$$(12.3) \quad \lambda_\epsilon^l \leq \nu^l \text{Max}_{v \in S_l} \frac{\int_{\Omega} v^2}{\int_{S_\epsilon} v^2},$$

where ν^l is the l th eigenvalue of the problem (6.5).

We prove now that there exists c_l such that

$$(12.4) \quad \text{Max}_{v \in S_l} \frac{\int_{\Omega} v^2}{\int_{S_\epsilon} v^2} \leq c_l \epsilon.$$

Suppose (12.4) is not true. Then we can extract a subsequence of ϵ (again denoted by ϵ) and $v_\epsilon \in S_l$ such that

$$(12.5) \quad \int_{\Omega} v_\epsilon^2 = 1,$$

$$(12.6) \quad \epsilon \int_{S_\epsilon} v_\epsilon^2 \rightarrow 0.$$

Since S_l is of finite dimension, the sequence $\{v_\epsilon\}$ remains bounded in $H_0^1(\Omega)$ also and it follows from (11.6) that

$$(12.7) \quad \int_{\Omega_\epsilon} v_\epsilon^2 \rightarrow 0.$$

If $v_\epsilon \rightarrow v_0$ in $L^2(\Omega)$ strong, then on the one hand (12.5) implies that

$$\int_{\Omega} v_0^2 = 1,$$

and on the other (12.7) implies

$$\int_{\Omega} v_0^2 = 0.$$

This contradiction proves (12.4) and hence the Proposition.

13. Asymptotic expansion

Taking into account the estimates we obtained in § 12, we normalize the eigenvector of the problem (10.1) as follows :

$$(13.1) \quad \int_{S_\epsilon} u_\epsilon^2 = \epsilon^{-1}.$$

Then we will have the following estimate :

$$(13.2) \quad \int_{\Omega_\epsilon} \nabla u_\epsilon \nabla u_\epsilon \leq c, \text{ independent of } \epsilon.$$

Considering all these estimates, we propose the following Ansatz to the problem (10.1).

$$(13.3) \quad u_\epsilon(x) = u_0(x, y) + \epsilon u_1(x, y) + \dots, y = x/\epsilon,$$

$$(13.4) \quad \lambda_\epsilon = \epsilon \lambda_1 + \epsilon^2 \lambda_2 + \dots$$

Each u_j is defined on $\Omega \times Y^*$ and it is Y periodic in y . We put the expressions (13.3) (13.4) in equation (10.1). We get

$$(13.5) \quad -(\epsilon^{-2} \Delta_y + 2\epsilon^{-1} \Delta_{\bullet y} + \Delta_\bullet)(u_0 + \epsilon u_1 + \dots) = 0,$$

$$(13.6) \quad \left(\epsilon^{-1} \frac{\partial}{\partial v_y} + v_j(y) \frac{\partial}{\partial x_j} \right) (u_0 + \epsilon u_1 + \dots) \\ = (\epsilon \lambda_1 + \epsilon^2 \lambda_2 + \dots)(u_0 + \epsilon u_1 + \dots)$$

where $\nu = (v_j)$ is the outer normal to S . For the notations used in this section see § 7. Now, we equate like powers of ϵ in the above relations and we solve the resulting equations. We obtain the following results :

$$(13.7) \quad u_0 \text{ is independent of } y : u_0(x, y) = u_0(x).$$

$$(13.8) \quad u_1(x, y) = -\chi^j(y) \frac{\partial u_0}{\partial x_j}(x) + \tilde{u}_1(x),$$

where χ^j are as defined below and $\tilde{u}_1$ is to be determined :

$$(13.9) \quad \begin{cases} a(\chi^j, v) = \int_S v_j v \, d\gamma \text{ for } v \in \mathcal{W}, \\ \chi^j \in \mathcal{W}. \end{cases}$$

We get the following equation satisfied by u_0 :

$$(13.10) \quad -r_{ij} \frac{\partial^2 u_0}{\partial x_i \partial x_j} = \lambda_1 u_0 \text{ in } \Omega,$$

with

$$(13.11) \quad r_{ij} = \frac{1}{|S|} a(\chi^i - y_i, \chi^j - y_j) \text{ for } i, j = 1, 2, \dots, N.$$

It follows from (13.11) that (r_{ij}) is a symmetric, positive definite matrix and so the following eigenvalue problem (called homogenized problem in this case) is well posed :

$$(13.12) \quad \begin{cases} \text{Find } (u, \lambda) \in H_0^1(\Omega) \times \mathbf{R} \text{ such that} \\ -r_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} = \lambda u \text{ in } \Omega, \\ u \neq 0. \end{cases}$$

The formal analysis so far shows that $\epsilon^{-1} \lambda_\epsilon \rightarrow$ to an eigenvalue of the problem (13.12). We prove this result in the later sections.

14. Results of homogenization

We denote by P_ϵ the prolongation operator constructed from the hole T_ϵ and P_ϵ satisfies (9.1). The following Theorem is analogous to Theorem (9.1).

Theorem (14.1). Let $l \geq 1$ be an integer. Let λ_ϵ^l be the l th eigenvalue of the problem (10.1) with u_ϵ^l be the corresponding eigenvector. Then $\{\epsilon^{-1} \lambda_\epsilon^l\}$ converges to an eigenvalue λ of the problem (13.12) and there exists a subsequence of ϵ (denoted by ϵ) such that

$$(14.1) \quad P_\epsilon u_\epsilon^l \rightarrow u \text{ in } H_0^1(\Omega) \text{ weak,}$$

where u is an eigenvector of the problem (13.12) corresponding to λ .

Proof.

Step 1. We set $\lambda_\epsilon = \lambda_\epsilon^l$, $u_\epsilon = u_\epsilon^l$ and

$$(14.2) \quad \xi_i^\epsilon = \frac{\partial u_\epsilon}{\partial x_i} \text{ in } \Omega_\epsilon \text{ for } i = 1, 2, \dots, N.$$

We put $\xi^\epsilon = 0$ in the holes. Because of the estimates (12.1) and (13.2), we can extract a subsequence of ϵ (denoted by ϵ again) such that

$$(14.3) \quad \xi_i^\epsilon \rightarrow \xi_i \text{ in } L^2(\Omega) \text{ weak for } i = 1, 2, \dots, N.$$

$$(14.4) \quad P_\epsilon u_\epsilon \rightarrow u \text{ in } H_0^1(\Omega) \text{ weak,}$$

$$(14.5) \quad \epsilon^{-1} \lambda_\epsilon \rightarrow \lambda.$$

Step 2. It follows from (10.1) that

$$(14.6) \quad \xi^\epsilon \cdot \nu_\epsilon = \lambda_\epsilon u_\epsilon \text{ on } S_\epsilon,$$

$$(14.7) \quad \operatorname{div} \xi^\epsilon = 0 \text{ in } \Omega_\epsilon.$$

We multiply equation (10.1) with $\eta \in D(\Omega)$:

$$\begin{aligned} \int_{\Omega_\epsilon} \nabla u_\epsilon \nabla \eta &= \lambda_\epsilon \int_{S_\epsilon} u_\epsilon \eta \\ &= \lambda_\epsilon \int_{\Omega_\epsilon} \frac{\partial \psi}{\partial y_i}(x/\epsilon) \frac{\partial}{\partial x_i} (u_\epsilon \eta) + c_0 \epsilon^{-1} \lambda_\epsilon \int_{\Omega_\epsilon} u_\epsilon \eta, \end{aligned}$$

and the passage to the limit will give us

$$\int_{\Omega} \xi \nabla \eta = c_0 \lambda \int_{\Omega} \theta u \eta,$$

which is equivalent to

$$(14.8) \quad -\operatorname{div} \xi = c_0 \lambda \theta u \text{ in } \Omega,$$

Step 3. We introduce w as solution to the following problem:

$$(14.9) \quad \begin{cases} a(w, v) = 0 \text{ for } v \in \mathcal{W} \\ w - \Pi \in \mathcal{W}, \end{cases}$$

where $\Pi(y)$ is a homogeneous polynomial of degree 1. We put

$$(14.10) \quad \eta = \nabla w \text{ in } Y^*.$$

Then η satisfies

$$(14.11) \quad \begin{cases} \operatorname{div} \eta = 0 \text{ in } Y^*, \\ \eta \cdot \nu = 0 \text{ on } S. \end{cases}$$

We extend η by zero in the hole and we put

$$(14.12) \quad \eta^\epsilon(x) = \eta(x/\epsilon) \text{ in } \Omega.$$

Then we will have

$$(14.13) \quad \operatorname{div} \eta^\epsilon = 0 \text{ in } \Omega.$$

Next we define

$$(14.14) \quad w^\epsilon(x) = \epsilon w(x/\epsilon).$$

As before (see Step 3, Theorem (9.1)), we get

$$(14.15) \quad w^\epsilon \rightarrow \Pi \text{ in } L^2(\Omega) \text{ strong.}$$

Step 4. Let $g \in D(\Omega)$. Multiply (14.7) by $g w^\epsilon$ and (14.13) by $g P_\epsilon u_\epsilon$ and subtract :

$$(14.16) \quad - \int_{\Omega_\epsilon} \xi_i^\epsilon \frac{\partial g}{\partial x_i} w^\epsilon + \int_{\Omega_\epsilon} \eta_i^\epsilon \frac{\partial g}{\partial x_i} P_\epsilon u_\epsilon + \int_{S_\epsilon} \lambda_\epsilon u_\epsilon g w^\epsilon = 0.$$

We can pass to the limit directly in the first two terms. For the third we use the formula (11.6). We get

$$(14.17) \quad - \int_{\Omega} \xi_i \frac{\partial g}{\partial x_i} \Pi + \int_{\Omega} m_Y \left(\chi \frac{\partial w}{\partial y_i} \right) \frac{\partial g}{\partial x_i} u + c_0 \lambda \theta \int_{\Omega} u g \Pi = 0$$

which gives

$$(14.18) \quad \xi_i \frac{\partial \Pi}{\partial x_i} = m_Y \left(\chi \frac{\partial w}{\partial y_i} \right) \frac{\partial u}{\partial x_i}.$$

Step 5. Now we take $\Pi = y_j$ ($j = 1, 2 \dots N$). The corresponding test function $w = -\chi^{(j)} + y_j$ where $\chi^{(j)}$ is defined by (13.9). So we get

$$(14.19) \quad \begin{aligned} \xi_j &= \frac{1}{|Y|} \int_{Y^*} \left(\delta_{ij} - \frac{\partial \chi^{(j)}}{\partial y_i} \right) \frac{\partial u}{\partial x_i}, \\ &= \frac{|S|}{|Y|} r_{ij} \frac{\partial u}{\partial x_i}. \end{aligned}$$

Combining this with (14.8) we obtain

$$(14.20) \quad - r_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} = \lambda u.$$

Step 6. To complete the proof, we have to show that $u \neq 0$. To see this, we take $g = u_\epsilon^2$ in (11.6) :

$$\epsilon \int_{S_\epsilon} u_\epsilon^2 = c_0 \int_{\Omega_\epsilon} u_\epsilon^2 + \epsilon \int_{\Omega_\epsilon} \frac{\partial \psi}{\partial y_i} (x/\epsilon) \frac{\partial}{\partial x_i} (u_\epsilon^2).$$

Using (13.1), we pass to the limit in this relation and we obtain

$$(14.21) \quad 1 = c_0 \theta \int_{\Omega} u^2,$$

from which it follows that $u \neq 0$.

We now give our final result in the Stekloff case.

Theorem (14.2)

Let $\{\lambda_\epsilon^l\}$, $\{u_\epsilon^l\}$ be the sequences of eigenvalues and eigenvectors of the problem (10.1). We suppose that the eigenvectors are normalized according to (13.1): $\int_{S_\epsilon} u_\epsilon^2 = \epsilon^{-1}$. Then

(i) $\lambda_\epsilon^l \rightarrow \lambda^l$, the l th eigenvalue of the problem (13.12).

(ii) there exists a subsequence of ϵ (again denoted by ϵ) such that

$$(14.22) \quad P_\epsilon u_\epsilon^l \rightarrow u^l \text{ in } H_0^1(\Omega) \text{ weak,}$$

where u^l satisfies (13.12) :

$$-r_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} = \lambda^l u \text{ in } \Omega, \quad u|_{\partial\Omega} = 0.$$

(iii) If λ^l is a simple eigenvalue of the problem (13.12) then given any eigenvector v^l associated to λ^l satisfying (14.2), we can choose an eigenvector v_ϵ^l of the problem (10.1) corresponding to λ_ϵ^l such that

$$(14.23) \quad \int_{S_\epsilon} v_\epsilon^{l^2} = \epsilon^{-1}$$

$$(14.24) \quad P_\epsilon v_\epsilon^l \rightarrow v^l \text{ in } H_0^1(\Omega) \text{ weak for the whole sequence.}$$

Proof. From the previous Theorem, we know that (ii) is true for a subsequence of ϵ . We obtain also the orthogonality condition for the eigenvectors by means of the method of Step 6 of Theorem (14.1) :

$$(14.25) \quad \int_{\Omega} u^l u^m = \frac{1}{c_0 \theta} \delta_{lm} \text{ for } l, m \geq 1.$$

Finally, it remains to prove that the limiting point λ^l of $\{\lambda_\epsilon^l\}$ is the l th eigenvalue of (13.12). For this, it suffices to verify that

$$(14.26) \quad \begin{cases} \text{(i) there is no eigenvalue other than } \{\lambda^l\}_{l=1}^{\infty} \text{ for the problem (13.12).} \\ \text{(ii) } \{u^l\}_{l=1}^{\infty} \text{ is an orthogonal basis in } L^2(\Omega). \end{cases}$$

In the following, we verify (14.26) (i) The proof of (14.26) (ii) is analogous.

Let us suppose that there is an eigenvalue μ different from λ^l with eigenvector u :

$$(14.27) \quad \begin{cases} -r_u \frac{\partial^2 u}{\partial x_i \partial x_i} = \mu u \text{ in } \Omega, \\ u \in H_0^1(\Omega), \\ \mu \neq \lambda^l \text{ for all } l, \\ \int_{\Omega} uu^l = 0 \text{ for each } l, \\ \int_{\Omega} u^2 = 1. \end{cases}$$

We can choose an integer l such that

$$(14.28) \quad \mu < \lambda^{l+1}.$$

We define w_ϵ as the solution of

$$(14.29) \quad \begin{cases} -\Delta w_\epsilon = 0 \text{ in } \Omega_\epsilon, \\ w_\epsilon \in W_\epsilon, \\ \frac{\partial w_\epsilon}{\partial \nu_\epsilon} = \epsilon \mu u \text{ on } S_\epsilon. \end{cases}$$

From the proof of Theorem (14.1), it is seen that

$$(14.30) \quad P_\epsilon w_\epsilon \rightarrow u \text{ in } H_0^1(\Omega) \text{ weak.}$$

Now consider the element

$$(14.31) \quad v_\epsilon = w_\epsilon - \epsilon \sum_{k=1}^l \left[\int_{S_\epsilon} w_\epsilon u_\epsilon^k \right] u_\epsilon^k.$$

Since

$$\int_{S_\epsilon} v_\epsilon u_\epsilon^k = 0 \text{ for } k = 1, 2, \dots, l,$$

we have

$$(14.32) \quad a^\epsilon(v_\epsilon, v_\epsilon) \geq \lambda_\epsilon^{l+1} \int_{S_\epsilon} v_\epsilon^2.$$

We now pass to the limit in (14.32) using (11.6). We have, first

$$\epsilon \int_{S_\epsilon} w_\epsilon u_\epsilon^k = \epsilon \int_{\Omega_\epsilon} \frac{\partial \psi}{\partial y_i}(x/\epsilon) \frac{\partial}{\partial x_i} (w_\epsilon u_\epsilon^k) + c_0 \int_{\Omega_\epsilon} w_\epsilon u_\epsilon^k,$$

and so

$$\epsilon \int_{S_\epsilon} w_\epsilon u_\epsilon^k \rightarrow c_0 \int_{\Omega} \theta u u^k = 0.$$

As a consequence,

$$(14.33) \quad P_\epsilon v_\epsilon \rightarrow u \text{ in } H_0^1(\Omega) \text{ weak.}$$

Using this and taking $g = \lambda_\epsilon^{l+1} v_\epsilon^2$ in (11.6), we get

$$(14.34) \quad \lambda_\epsilon^{l+1} \int_{S_\epsilon} v_\epsilon^2 \rightarrow c_0 \lambda^{l+1} \theta.$$

By similar arguments, we prove that

$$(14.35) \quad a^\epsilon(v_\epsilon, v_\epsilon) \rightarrow c_0 \theta \mu.$$

Now passing to the limit in (14.32) we get

$$(14.36) \quad \mu \geq \lambda^{l+1}$$

which contradicts (14.28).

To prove (iii), we remark that λ_ϵ^l is simple for sufficiently small ϵ if λ^l is so. Afterwards, it suffices to pick up an eigenvector v_ϵ^l associated to λ_ϵ^l such that

$$(14.37) \quad \int_{S_\epsilon} v_\epsilon^l v^l > 0.$$

Part C: Neumann eigenvalue problem

15. Problem to be studied

We consider the following eigenvalue problem

$$(15.1) \quad \begin{cases} \text{Find } (u_\epsilon, \lambda_\epsilon) \in W_\epsilon \times \mathbb{R} \text{ such that} \\ -\Delta u_\epsilon = \lambda_\epsilon u_\epsilon \text{ in } \Omega_\epsilon, \\ \frac{\partial u_\epsilon}{\partial \nu_\epsilon} = 0 \text{ on } S_\epsilon, \\ \int_{\Omega_\epsilon} u_\epsilon^2 = 1, \end{cases}$$

where we have used the notations of § 2 and § 10. Let us denote by $\{\lambda_\epsilon^l\}$ and $\{u_\epsilon^l\}$ the sequences of eigenvalues and the corresponding eigenvectors of the problem (15.1). We know that

$$(15.2) \quad \begin{cases} \lambda_\epsilon^1 < \lambda_\epsilon^2 \leq \lambda_\epsilon^3 \dots \rightarrow \infty, \\ \{u_\epsilon^l\} \text{ form an orthonormal basis in } L^2(\Omega_\epsilon). \end{cases}$$

Here, we study the asymptotic behaviour of λ_ϵ^l and u_ϵ^l as $\epsilon \rightarrow 0$. We prove that $\{\lambda_\epsilon^l\}$ converges to the l th eigenvalue of the "homogenized problem".

Since the method followed here is similar to Stekloff case, we do not give details of the proofs.

16. Estimates on the eigenvalues

Proposition (16.1). Let λ_ϵ^l be the l th eigenvalue of the problem (15.1). Then there exists a constant c_l independent of ϵ such that

$$\lambda_\epsilon^l \leq c_l.$$

Proof. As in the proof of Proposition (12.1), by the use of mini-max principle we get

$$(16.1) \quad \lambda_\epsilon^1 \leq v^1 \max_{v \in S_1} \frac{\int_{\Omega} v^2}{\int_{\Omega} v^2}.$$

But we saw in the course of the proof of Proposition (12.1) that the right side of (16.1) is bounded above independently of ϵ and this completes the proof.

17. Asymptotic analysis

With the help of the estimates in §16 and the normalization condition for the eigenvectors, we propose the following Ansatz for the problem (15.1) :

$$(17.1) \quad u_\epsilon(x) = u_0(x, y) + \epsilon u_1(x, y) + \dots, \quad y = x/\epsilon,$$

$$(17.2) \quad \lambda_\epsilon = \lambda_0 + \epsilon \lambda_1 + \dots$$

where u_j is defined on $\Omega \times Y^*$ and it is Y periodic in y .

We substitute these expressions in equation (15.1) :

$$(17.3) \quad -(\epsilon^{-2} \Delta_y + 2\epsilon^{-1} \Delta_{xy} + \Delta_x)(u_0 + \epsilon u_1 + \dots) \\ = (\lambda_0 + \epsilon \lambda_1 + \dots)(u_0 + \epsilon u_1 + \dots),$$

$$(17.4) \quad \left(\epsilon^{-1} \frac{\partial}{\partial v_y} + v_j \frac{\partial}{\partial x_j} \right) (u_0 + \epsilon u_1 + \dots) = 0.$$

As before, we equate the powers of ϵ on either side and we obtain the following results :

$$(17.5) \quad u_0 \text{ is independent of } y : u_0(x, y) = u_0(x).$$

$$(17.6) \quad u_1(x, y) = -\chi^j(y) \frac{\partial u_0}{\partial x_j}(x) + \tilde{u}_1(x),$$

where χ^j are defined in (13.9). We get the following equation satisfied by u_0 and λ_0 :

$$(17.7) \quad -s_{ij} \frac{\partial^2 u_0}{\partial x_i \partial x_j} = \lambda_0 u_0 \text{ in } \Omega,$$

with

$$(17.8) \quad s_{ij} = \frac{1}{|Y^*|} a(\chi^i - y_i, \chi^j - y_j) \text{ for all } i, j.$$

So, the corresponding "homogenized problem" in this case can be formulated as follows :

$$(17.9) \quad \begin{cases} \text{Find } (u, \lambda) \in H_0^1(\Omega) \times \mathbf{R} \text{ such that} \\ -s_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} = \lambda u \text{ in } \Omega \\ u \neq 0. \end{cases}$$

In the following section, we justify the formal calculations done above.

18. Homogenization theorem

The proof of the following result proceeds along the lines of Theorems (14.1) and (14.2) and therefore omitted.

Theorem (18.1). Let $\{\lambda_\epsilon^l\}$, $\{u_\epsilon^l\}$ be the sequences of eigenvalues and the corresponding eigenvectors of the problem (15.1) satisfying (15.2). Then

(i) $\lambda_\epsilon^l \rightarrow \lambda^l$, the l th eigenvalue of the problem (17.9).

(ii) there exists a subsequence of ϵ (denoted again by ϵ) such that

$$P_\epsilon u_\epsilon^l \rightarrow u^l \text{ in } H_0^1(\Omega) \text{ weak,}$$

where u^l is an eigenvector corresponding to λ^l .

(iii) If λ^l is simple, then given any eigenvector v^l corresponding to λ^l such that

$$(18.1) \quad \int_{\Omega} v^{l^2} = 1/\theta,$$

we can choose an eigenvector v_ϵ^l of the problem (15.1) corresponding to λ_ϵ^l such that $P_\epsilon v_\epsilon^l \rightarrow v^l$ in $H_0^1(\Omega)$ weak for the whole sequence of ϵ .

19. Correctors and error estimates

In this paragraph, we briefly mention a method of finding correctors of first order for simple eigenvalue of the Neumann problem. We will also see that the problem of error estimates is reduced to that of stationary problem for which we refer to Lions [13]. For the study of correctors in the homogenization theory, we refer to Bensoussan *et al* [3], Bourgat and Dervieux [5], Kesavan [12].

We consider a simple eigenvalue λ_0 of the problem (17.9) with eigenvector u_0 satisfying

$$(19.1) \quad \int_{\Omega} u_0^2 = 1/\theta.$$

Let u_ϵ , λ_ϵ be eigenvector and eigenvalue of the problem (15.1) satisfying

$$(19.2) \quad P_\epsilon u_\epsilon \rightarrow u_0 \text{ in } H_0^1(\Omega) \text{ weak for the whole sequence of } \epsilon,$$

$$(19.3) \quad \lambda_\epsilon \rightarrow \lambda_0.$$

Let us define w_ϵ as the solution of

$$(19.4) \quad \begin{cases} -\Delta w_\epsilon = \lambda_0 u_0 \text{ in } \Omega_\epsilon, \\ w_\epsilon \in W_\epsilon, \\ \frac{\partial w_\epsilon}{\partial \nu_\epsilon} = 0 \text{ on } S_\epsilon. \end{cases}$$

The proof of Theorem (18.1) shows that

$$(19.5) \quad P_\epsilon w_\epsilon \rightarrow u_0 \text{ in } H_0^1(\Omega) \text{ weak.}$$

Multiplying (19.4) by u_ϵ and (15.1) by w_ϵ , we obtain

$$(19.6) \quad \lambda_\epsilon (u_\epsilon, w_\epsilon)_{\Omega_\epsilon} = \lambda_0 (u_0, u_\epsilon)_{\Omega_\epsilon},$$

where $(\cdot, \cdot)_{\Omega_\epsilon}$ denotes the scalar product in $L^2(\Omega_\epsilon)$.

First let us prove the following:

Theorem (19.1)

With $\lambda_0, u_0, \lambda_\epsilon, u_\epsilon, w_\epsilon$ defined as above, we have

$$(19.7) \quad |\lambda_\epsilon - \lambda_0| \leq c \|w_\epsilon - u_0\|_{L^2(\Omega_\epsilon)},$$

for sufficiently small ϵ and where $c > 0$ is a constant independent of ϵ .

Proof. We have

$$\lambda_\epsilon - \lambda_0 = \frac{\lambda_0}{(u_\epsilon, w_\epsilon)_{\Omega_\epsilon}} (u_\epsilon, u_0 - w_\epsilon)_{\Omega_\epsilon}.$$

But $(u_\epsilon, w_\epsilon)_{\Omega_\epsilon} \rightarrow 1$ as $\epsilon \rightarrow 0$ and so we obtain (19.7).

Now, we give a result which estimates the error between the eigenvectors by choosing an eigenvector corresponding to λ_ϵ . For that, we first define z_ϵ uniquely as follows :

$$(19.8) \quad \begin{cases} a^\epsilon(z_\epsilon, v) - \lambda_\epsilon(z_\epsilon, v)_{\Omega_\epsilon} = \lambda_\epsilon(u_\epsilon, v)_{\Omega_\epsilon} - \lambda_0(u_0, v)_{\Omega_\epsilon} \text{ for } v \in W_\epsilon, \\ z_\epsilon \in W_\epsilon, \\ (z_\epsilon, u_\epsilon)_{\Omega_\epsilon} = 0. \end{cases}$$

Then

$$(19.9) \quad \tilde{u}_\epsilon = z_\epsilon + w_\epsilon,$$

$\tilde{u}_\epsilon$ is an eigenvector of the problem (15.1) corresponding to λ_ϵ . We have the following estimate for z_ϵ :

Lemma (19.1). There is a constant $c > 0$ independent of ϵ such that

$$(19.10) \quad \|z_\epsilon\|_{1, \Omega_\epsilon} \leq c \|\lambda_\epsilon w_\epsilon - \lambda_0 u_0\|_{L^2(\Omega_\epsilon)}.$$

Proof. Suppose (19.10) is not true. Then there will be a subsequence of ϵ (again denoted by ϵ) such that

$$(19.11) \quad \|z_\epsilon\|_{1, \Omega_\epsilon} = 1,$$

$$(19.12) \quad \|\lambda_\epsilon w_\epsilon - \lambda_0 u_0\|_{L^2(\Omega_\epsilon)} \rightarrow 0.$$

We can pass to the limit in (19.8) as we did in equation (15.1). We obtain

$$(19.13) \quad P_\epsilon z_\epsilon \rightarrow z_0 \text{ in } H_0^1(\Omega) \text{ weak,}$$

where z_0 satisfies

$$(19.14) \quad \begin{cases} -s_{ij} \frac{\partial^2 z_0}{\partial x_i \partial x_j} = \lambda_0 z_0 \text{ in } \Omega, \\ \int_\Omega z_0 u_0 = 0. \end{cases}$$

Since λ_0 is simple, this implies that

$$(19.15) \quad z_0 = 0.$$

On the other hand, taking $v = z_\epsilon$ in (19.18), we get

$$1 = \lambda_\epsilon (z_\epsilon, z_\epsilon)_{\Omega_\epsilon} + (\lambda_\epsilon w_\epsilon - \lambda_0 u_0, z_\epsilon)_{\Omega_\epsilon},$$

which, at the limit, gives a contradiction.

Theorem (19.2)

Let $\lambda_0, \lambda_\epsilon, w_\epsilon, u_0, \tilde{u}_\epsilon$ be defined as above. Then there is a constant $c > 0$ independent of ϵ such that

$$(19.16) \quad \|\tilde{u}_\epsilon - u_0\|_{L^2(\Omega_\epsilon)} \leq c \|w_\epsilon - u_0\|_{L^2(\Omega_\epsilon)}.$$

Proof. We write

$$\tilde{u}_\epsilon - u_0 = z_\epsilon + w_\epsilon - u_0.$$

The proof is completed by using (19.11) and Theorem (19.1) and the fact that

$$\|z_\epsilon\|_{L^2(\Omega_\epsilon)} \leq c \|z_\epsilon\|_{1, (\Omega_\epsilon)}.$$

Now we give a first order corrector for the eigenvalue λ_ϵ . We use (19.6). We express

$$\lambda_\epsilon = \lambda_0 + \epsilon \lambda_1 + \dots$$

$$u_\epsilon = u_0 + \epsilon u_1 + \dots$$

$$w_\epsilon = u_0 + \epsilon w_1 + \dots$$

We put these expressions in (19.6) and identify powers of ϵ : we get

$$(19.17) \quad \lambda_1 = -\lambda_0 \frac{(u_0, w_1)_{\Omega_\epsilon}}{(u_0, u_0)_{\Omega_\epsilon}},$$

where w_1 is the first order corrector for w_ϵ . The estimate we give below is better than that given by Theorem (19.1).

Theorem (19.3)

Let $\lambda_\epsilon, \lambda_0, w_\epsilon, u_0$ be as in the previous Theorem. Let λ_1 be [defined by (19.17). Then we have

$$\begin{aligned} |\lambda_\epsilon - \lambda_0 - \epsilon \lambda_1| &\leq c \{ \epsilon \|w_\epsilon - u_0\|_{L^2(\Omega_\epsilon)} + \|w_\epsilon - u_0\|_{L^2(\Omega_\epsilon)}^2 \\ &\quad + \|w_\epsilon - u_0 - \epsilon w_1\|_{L^2(\Omega_\epsilon)} \}, \end{aligned}$$

where c is a constant independent of ϵ .

Proof. We have

$$\begin{aligned} \lambda_\epsilon - \lambda_0 - \epsilon \lambda_1 &= \frac{\lambda_0}{(u_0, u_0)_{\Omega_\epsilon}} (u_0, u_0 + \epsilon w_1 - w_\epsilon)_{\Omega_\epsilon} \\ &\quad - (\lambda_\epsilon - \lambda_0) \frac{(u_0, \epsilon w_1)_{\Omega_\epsilon}}{(u_0, u_0)_{\Omega_\epsilon}} + \left[\frac{\lambda_\epsilon}{(u_0, u_0)_{\Omega_\epsilon}} (u_0, w_\epsilon)_{\Omega_\epsilon} - \lambda_0 \right]. \end{aligned}$$

We only have to estimate the last term. But using (19.6), one can express the last term as follows:

$$\begin{aligned} \frac{\lambda_\epsilon}{(u_0, u_0)_{\Omega_\epsilon}} (u_0, w_\epsilon)_{\Omega_\epsilon} - \lambda_0 &= \frac{\lambda_\epsilon}{(u_0, u_0)_{\Omega_\epsilon}} (u_0 - \tilde{u}_\epsilon, w_\epsilon - u_0)_{\Omega_\epsilon} \\ &\quad + \frac{\lambda_\epsilon - \lambda_0}{(u_0, u_0)_{\Omega_\epsilon}} (u_0, u_0 - \tilde{u}_\epsilon) \leq c \|w_\epsilon - u_0\|_{L^2(\Omega_\epsilon)}^2, \end{aligned}$$

by Theorems (19.1) and (19.2). This completes the proof.

Acknowledgements

The author is thankful to Professor J L Lions, College de France, Paris, for his help and encouragement in carrying out this work during his stay at Paris. The discussions, the author had with Professor L Tartar, Orsay, France, were very fruitful.

The author thanks the referee for having brought Rauch's article "The Mathematical Theory of Crushed Ice", Partial Differential Equations and Related topics, Springer Verlag, *Lecture Notes in Mathematics*, No. 446, 1975, to his notice. In this article, Rauch considers what he calls the crushed ice problem; obviously this is related to our Dirichlet problem. Rauch shows, in dimension 3, that the first eigenvalue $\lambda_1 \rightarrow \infty$ when $n \rightarrow \infty$, $r \rightarrow 0$, $nr^3 = \text{constant}$ and $nr^2 \rightarrow \infty$ where n is the number of holes and r is the size of each hole. In our case, $n = 0$ (ϵ^{-3}), $r = 0$ (ϵ) and so $nr^2 = 0$ (ϵ^{-1}) $\rightarrow \infty$ as $\epsilon \rightarrow 0$. Therefore, it is not surprising that $\lambda_1 \rightarrow \infty$ for the Dirichlet problem. However, our method even gives an asymptotic expansion valid not only for the first eigenvalue but also for all eigenvalues.

References

- [1] Baouendi M S 1967 *Bull. Soc. Math. de France* **95** 45
- [2] Baouendi M S and Goulaouic C 1968 *C.R. Acad. Sci. Serie A* tome **266** 336
- [3] Bensoussan A, Lions J L and Papanicolaou G 1978 *Asymptotic methods in periodic structures* (Amsterdam: North Holland)
- [4] Bergmann S and Schiffer M 1953 *Kernel functions and elliptic differential equations in mathematical physics* (New York: Academic Press)
- [5] Bourgat J F and Dervieux A 1978 Etude des correcteurs provenant du developpement asymptotique, Rapport Laboria No. 278

- [6] Bramble J H and Osborn J E 1972 in *The mathematical foundations of the finite element method with applications to the partial differential equations* (ed. : A K Aziz) (New York : Academic Press)
- [7] Cioranescu D 1978 Sur quelques equations aux derivees partielles posees par la Mechanique des milieux continus, These d'Etat Paris.
- [8] Cioranescu D and Saint Jean Paulin J 1978 Homogeneisation dans des ouverts avec cavite's reparties periodiquement, Rapport de l'universite P et M Curie, Paris
- [9] Duvaut G 1977 *Comportement macroscopique d'une plaque perforee periodiquement*, Lecture Notes (Berlin : Springer-Verlag) 594
- [10] Geymonat G and Grisvard P 1967 *Le Matematiche* 22 2
- [11] Goudjo C 1970 Problems aux limites dans les espaces de Sobolov avec poids These de 3^{eme} cycle Nice
- [12] Kesavan S 1979 Approximation des problemes de valens propres lineaires et non-lineaires These d'Etat, Paris
- [13] Lions J L 1975-77 Cours d'homogeneisation, College de France
- [14] Tartar L 1977 Problems d'homogeneisation dans les equations aux derivees partielles, College de France Cours Peccot
- [15] Vanninathan M 1978 *C.R. Acad. Sci. Paris Serie A* tome 287 p. 823
- [16] Vanninathan M 1978 *C.R. Acad. Sci. Serie A* tome 287 p. 403
- [17] Weinstein A and Steinger W 1972 *Methods of intermediate problems for eigenvalues. Theory and ramifications* (New York : Academic Press)

A note on the mean value of L-series

V V RANE

Scuola Normale Superiore, Piazza dei Cavalieri, 7 Pisa 56100, Italy

MS received 10 December 1979 ; revised 21 March 1981

Abstract. Using Hilbert's inequality, we give a new asymptotic formula (uniform in q and T) for

$$\sum_{\substack{\chi \pmod{q} \\ \chi \text{ primitive}}} \int_T^{2T} |L(\tfrac{1}{2} + it, \chi)|^4 dt$$

Keywords. L-series; primitive characters; mean value.

1. Introduction

The object of this paper is to give a new asymptotic formula for

$$\sum_{\chi}^* \int_T^{2T} |L(\tfrac{1}{2} + it, \chi)|^4 dt,$$

uniform in q and T where $\sum_{\chi}^*$ denotes summation over primitive characters modulo q .

Before stating our theorem, we introduce our notation.

N denotes the number of primitive characters $(\bmod q)$. $\sum_{\chi}^*$ denotes summation over primitive characters $(\bmod q)$. We assume $N \geq 1$. It should be noted that $N \geq 1$, if and only if either q is odd or 4 divides q . We write $q = p_1^{l_1} p_2^{l_2} \dots p_r^{l_r}$, where p_i 's are primes and r = number of distinct prime factors of q .

$$\text{Also } N = \prod_{i=1}^r \theta(p_i^{l_i}),$$

where $\theta(p_i^{l_i}) = p_i - 2$ for $l_i = 1$

$$= p_i^{l_i-2} (p_i - 1)^2 \text{ for } l_i \geq 2,$$

for $i = 1, 2, \dots, r$.

In what follows, $\phi(n)$, $d(n)$, $\mu(n)$ denote Euler's totient function, divisor function and Möbius function respectively.

We note that

$$\frac{q}{N} \ll \frac{q^2}{\phi^2(q)} \ll (\log \log 3q)^2.$$

We use the fact

$$\sum_{n \leq y} d^2(n) \ll y \log^3 y \text{ for } y > 1,$$

quite frequently in the sequel. For convenience, we write $X = qT$. In the sequel, p with or without subscript denotes a prime number. Our theorem heavily depends on lemma 5, which, in its turn, is a consequence of lemmas 3 and 4.

2. Statement of the theorem

Let $L(s, \chi)$ be Dirichlet's L -function to the modulus q . Let $X = qT$ and $N =$ the number of primitive characters $(\text{mod } q)$ and let $\gamma =$ distinct prime factors of q . Then for $T \geq 2$ and for all q (for which $N \geq 1$),

$$\begin{aligned} & \frac{1}{N} \sum_{\chi}^* \int_T^{2T} |L(\tfrac{1}{2} + it, \chi)|^4 dt \\ &= \frac{1}{2\pi^2} \frac{\prod_{p|q} \left(1 - \frac{1}{p}\right)}{\prod_{p|q} \left(1 - \frac{1}{p^2}\right)} T \log^4 X + O(2^{\gamma} T \log^3 X (\log \log 3q)^5). \end{aligned}$$

Remark. The special case $q = 1$ gives the theorem of Ingham for zeta-function viz.,

$$\int_T^{2T} |\zeta(\tfrac{1}{2} + it)|^4 dt = \frac{T}{2\pi^2} \log^4 T + O(T \log^3 T).$$

3. Proof of the theorem

For every primitive character $\chi(\text{mod } q)$, we have functional equation of the type

$$L(s, \chi) = \psi(s, \chi) L(1 - s, \bar{\chi}).$$

We note that

$$(qt)^{1-\sigma} \ll \psi(s, \chi) \ll (qt)^{1-\sigma}$$

where $s = \sigma + it$, uniformly for all primitive characters $\chi(\text{mod } q)$ and for σ in a fixed interval.

Lemma 1. (a) : Let

$$s = \sigma + it, |\sigma - \frac{1}{2}| \leq \frac{1}{100}, T \leq t \leq 2T, w = u + iv.$$

Then, for any primitive character $\chi \pmod{q}$,

$$\begin{aligned} L^2(s, \chi) &= \sum_{n=1}^{\infty} d(n) \chi(n) e^{-n/X} n^{-s} \\ &\quad + \psi^2(s, \chi) \sum_{n \leq X} d(n) \overline{\chi(n)} n^{s-1} \\ &\quad - \frac{1}{2\pi i} \int_{u=-\frac{3}{4}} \psi^2(s+w, \chi) \left(\sum_{n>X} d(n) \overline{\chi(n)} n^{w+s-1} \right) \Gamma(w) X^w dw \\ &\quad - \frac{1}{2\pi i} \int_{u=\frac{1}{4}} \psi^2(s+w, \chi) \left(\sum_{n \leq X} d(n) \overline{\chi(n)} n^{w+s-1} \right) \Gamma(w) X^w dw \\ &\quad + O(T^{-10}). \end{aligned}$$

Remark. Note that

$$L^2(s, \chi) = \sum_{n=1}^{\infty} \frac{d(n) \chi(n)}{n^s} \text{ for } \sigma > 1.$$

This lemma can be proved by starting with the integral

$$\frac{1}{2\pi i} \int_{u=2} L^2(s+w, \chi) \Gamma(w) X^w dw;$$

moving the line of integration to $\text{Re } w = u = -3/4$; using the functional equation for $L^2(s+w, \chi)$ there, splitting the series for $L^2(1-s-w, \bar{\chi})$ and consequently, the integral in an obvious manner and moving the line of integration in one of them back to $u = 1/4$. Also, we note that the error term $O(T^{-10})$ occurs, only when $q = 1$ i.e. in the case of ζ -function, since in that case, before effecting the change of line of integration for

$$\frac{1}{2\pi i} \int_{u=2} \zeta^2(s+w) \Gamma(w) X^w dw$$

we have to break the line of integration suitably, to avoid the pole of zeta function. We rewrite this lemma in a modified form to suit our needs as follows :

Lemma 1. (b): For $|\sigma - \frac{1}{2}| \leq \frac{1}{100}$ and $T \leq t \leq 2T$,

$$\begin{aligned} \psi(s, \chi)^{-1} L^2(s, \chi) &= \psi(s, \chi)^{-1} \sum_{n \leq X} d(n) \chi(n) n^{-s} \\ &\quad + \psi(s, \chi) \sum_{n \leq X} d(n) \overline{\chi(n)} n^{s-1} + \psi(s, \chi)^{-1} \sum_{n \leq X} d(n) \chi(n) n^{-s} (e^{-n/X} - 1) \\ &\quad + \psi(s, \chi)^{-1} \sum_{n > X} d(n) \chi(n) n^{-s} \exp(-n/X) \end{aligned}$$

$$\begin{aligned}
& - \frac{\psi(s, \chi)^{-1}}{2\pi i} \int_{u=-\frac{3}{2}} \psi^2(s+w, \chi) \left(\sum_{n > X} d(n) \overline{\chi(n)} n^{w+s-1} \right) \Gamma(w) X^w dw \\
& - \frac{\psi(s, \chi)^{-1}}{2\pi i} \int_{u=\frac{1}{2}} \psi^2(s+w, \chi) \left(\sum_{n \leq X} d(n) \overline{\chi(n)} n^{w+s-1} \right) \Gamma(w) X^w dw \\
& + O(T^{-10}) = \sum_{k=1}^7 J_k(s, \chi), \text{ say.}
\end{aligned}$$

We write $J_k(\frac{1}{2} + it, \chi) = J_k(\chi)$ and since $J_1(\chi), J_2(\chi)$ are complex conjugates, we get the following :

Lemma 2. We have

$$\begin{aligned}
& = \frac{1}{N} \sum_{\chi}^* \int_T^{2T} \left| L\left(\frac{1}{2} + it, \chi\right) \right|^4 dt \\
& = \frac{1}{N} \sum_{\chi}^* \int_T^{2T} (J_1(\chi) + J_2(\chi))^2 dt \\
& \quad + O\left(\frac{1}{N} \sum_{\chi}^* \int_T^{2T} \sum_{k=3}^6 |J_k^2(\chi)| dt\right) \\
& \quad + O\left\{\frac{1}{N} \sum_{\chi}^* \int_T^{2T} (J_1(\chi) + J_2(\chi)) \left(\sum_{k=3}^6 J_k(\chi)\right) dt\right\} \\
& = \frac{2}{N} \sum_{\chi}^* \int_T^{2T} |J_1^2(\chi)| dt \\
& \quad + O\left(\frac{1}{N} \sum_{\chi}^* \int_T^{2T} (J_1^2(\chi) + J_2^2(\chi)) dt\right) \\
& \quad + O\left(\frac{1}{N} \sum_{\chi}^* \int_T^{2T} \sum_{k=3}^6 |J_k^2(\chi)| dt\right) \\
& \quad + O\left(\frac{1}{N} \sum_{\chi}^* \int_T^{2T} (J_1(\chi) + J_2(\chi)) \left(\sum_{k=3}^6 J_k(\chi)\right) dt\right).
\end{aligned}$$

Lemma 3. Let $\{b(n)\}$ $n = 1, 2, \dots, R$ be a sequence of complex numbers and Q be any integer. Then

$$\begin{aligned}
& \sum_{\substack{m \neq n \\ m \equiv n \pmod{Q}}} \sum \frac{b(m) \overline{b(n)}}{\log m - \log n} \\
& = O\left(\sum_{n \leq Q/4} \frac{|b(n)|^2}{\log(Q/n)}\right) + O\left(\frac{1}{Q} \sum_{n=1}^R |nb(n)|^2\right).
\end{aligned}$$

Proof. We have

$$\begin{aligned}
& \sum_{\substack{m \neq n \\ m \equiv n \pmod{Q}}} \sum \frac{b(m) \overline{b(n)}}{\log m - \log n} \\
& = \sum_{j=1}^Q \left(\sum_{\substack{m' \neq n' \\ m', n' \geq 0}} \frac{b(m'Q+j) \overline{b(n'Q+j)}}{\log(m'Q+j) - \log(n'Q+j)} \right).
\end{aligned}$$

Now according to equation 1.7 of Montgomery and Vaughan [1], if $\{\lambda_n\}$ $n = 1, 2, \dots, R$, is a sequence of distinct real numbers and if

$$\delta_n = \min_{m \neq n} |\lambda_n - \lambda_m|,$$

then for any sequence $\{c(n)\}$ $n = 1, 2, \dots, R$ of complex numbers, we have

$$\left| \sum_{m \neq n} \sum \frac{c(m) \overline{c(n)}}{\lambda_m - \lambda_n} \right| \leq \frac{3\pi}{2} \sum_n |c(n)|^2 \delta_n^{-1}.$$

Using this result, we get

$$\sum_{\substack{m' \neq n' \\ m', n' \geq 0}} \sum \frac{b(m'Q+j) \overline{b(n'Q+j)}}{\log \frac{m'Q+j}{n'Q+j}} \ll \sum_{n' \geq 0} |b(n'Q+j)|^2 \delta_{n'}^{-1},$$

where $\delta_{n'} = \log \frac{(n'+1)Q+j}{n'Q+j},$

and for $n' \geq 1,$

$$\delta_{n'} \gg \frac{1}{n'} \text{ for } 1 \leq j \leq Q.$$

Thus

$$\begin{aligned} \sum_{\substack{m' \neq n' \\ m', n' \geq 0}} \sum \frac{b(m'Q+j) \overline{b(n'Q+j)}}{\log \frac{m'Q+j}{n'Q+j}} &\ll \frac{|b(j)|^2}{\log \frac{Q+j}{j}} \\ &+ \sum_{n' \geq 1} n' |b(n'Q+j)|^2 \\ &\ll \frac{|b(j)|^2}{\log \frac{Q+j}{j}} + \frac{1}{Q} \sum_{n' \geq 1} (n'Q+j) |b(n'Q+j)|^2. \end{aligned}$$

Therefore,

$$\begin{aligned} \sum_{\substack{m \neq n \\ m \equiv n \pmod{Q}}} \sum \frac{b(m) \overline{b(n)}}{\log(m/n)} \\ &\ll \sum_{j=1}^Q \frac{|b(j)|^2}{\log \frac{Q+j}{j}} + \frac{1}{Q} \sum_n n |b(n)|^2 \\ &\ll \sum_{j \leq Q^{1/4}} \frac{|b(j)|^2}{\log \frac{Q+j}{j}} + \frac{1}{Q} \sum_n n |b^2(n)| \end{aligned}$$

Thus

$$\begin{aligned} \sum_{\substack{m \neq n \\ m \equiv n \pmod{Q}}} \sum \frac{b(m) \overline{b(n)}}{\log(m/n)} \\ &\ll \sum_{n \leq Q^{1/4}} \frac{|b^2(n)|}{\log(Q/n)} + \frac{1}{Q} \sum_n n |b^2(n)|. \end{aligned}$$

Lemma 4. For any sequence $\{b(n)\}$ of complex numbers, we have,

$$\begin{aligned} & \sum_{\chi \pmod{q}}^* \sum_{m \neq n} \frac{b(n) \overline{b(m)} \chi(n) \overline{\chi(m)}}{\log(m/n)} \\ & \ll \sum_{k|q} |\mu(q/k)| \left\{ \phi(k) \sum_{n \leq k/4} \frac{|b^2(n)|}{\log(k/n)} \right. \\ & \quad \left. + \frac{\phi(k)}{k} \sum_n |b^2(n)| \right\} \\ & \ll 2^r \left\{ \max_{k|q} \left(k \sum_{n \leq k/4} \frac{|b^2(n)|}{\log(k/n)} \right) + \sum_n |b^2(n)| \right\}. \end{aligned}$$

Proof. We have,

$$\begin{aligned} & \sum_{\chi \pmod{q}}^* \sum_{m \neq n} \frac{b(n) \overline{b(m)} \chi(n) \overline{\chi(m)}}{\log(m/n)} \\ & = \sum_{\substack{m \neq n \\ (m, n) = 1 \pmod{q}}} \sum \frac{b(n) \overline{b(m)}}{\log m/n} \sum_{\chi \pmod{q}}^* \chi(n) \overline{\chi(m)}. \end{aligned}$$

Now, by virtue of well-known properties of primitive characters,

$$\begin{aligned} & \sum_{\chi \pmod{q}}^* \chi(n) \overline{\chi(m)} \\ & = \prod_{j=1}^r \left(\sum_{\chi \pmod{p_j^{t_j}}}^* \chi(n) \overline{\chi(m)} \right) \\ & = \prod_{j=1}^r \left(\sum_{t_j=0}^{t_j} \mu(p_j^{t_j-t_j}) \sum_{\chi \pmod{p_j^{t_j}}} \chi(n) \overline{\chi(m)} \right) \\ & = \sum_{k|q} \mu(q/k) \left(\sum_{\chi \pmod{k}} \chi(n) \overline{\chi(m)} \right). \end{aligned}$$

In view of the relation, $\sum_{\chi \pmod{Q}} \chi(n) \overline{\chi(m)} = \phi(Q)$ if $m \equiv n \pmod{Q}$, $(n, Q) = 1$; and zero, otherwise, for any modulus Q , we get

$$\begin{aligned} & \sum_{\chi \pmod{q}}^* \sum_{m \neq n} \frac{b(n) \overline{b(m)} \chi(n) \overline{\chi(m)}}{\log(m/n)} \\ & = \sum_{k|q} \phi(k) \mu(q/k) \left(\sum_{\substack{m \neq n \\ m \equiv n \pmod{k} \\ (n, k) = 1}} \frac{b(n) \overline{b(m)}}{\log(m/n)} \right). \end{aligned}$$

The proof now follows by appealing to the preceding lemma.

Lemma 5. We have for any sequence $\{a(n)\}$ of complex numbers,

$$\begin{aligned} & \frac{1}{N} \sum_{\chi}^* \int_T^{T+H} \left| \sum_{n \leq X} a(n) \chi(n) n^{-it} \right|^2 dt \\ &= H \sum_{\substack{n \leq X \\ (n, q)=1}} |a^2(n)| + O\left(\frac{2^r}{N} \max_{k|q} k \sum_{n \leq k/4} \frac{|a^2(n)|}{\log(k/n)}\right) \\ &+ O\left(\frac{2^r}{N} \sum_{n \leq X} n |a^2(n)|\right). \end{aligned}$$

Lemma 6. We have

$$\begin{aligned} & \frac{2}{N} \sum_{\chi}^* \int_T^{2T} |J_1^2(\chi)| dt = 2T \sum_{\substack{n \leq X \\ (n, q)=1}} \frac{d^2(n)}{n} \\ &+ O\left(2^r \frac{q}{N} (T \log^3 X) (\log \log 3q)\right) \\ &= \frac{1}{2\pi^2} \prod_{p|q} \left(\frac{1 - \frac{1}{p}}{1 - \frac{1}{p^2}} \right)^4 T (\log X)^4 + O(2^r T \log^3 X (\log \log 3q)^5). \end{aligned}$$

Proof. Using the preceding lemma,

$$\begin{aligned} & \frac{2}{N} \sum_{\chi}^* \int_T^{2T} |J_1^2(\chi)| dt = 2T \sum_{\substack{n \leq X \\ (n, q)=1}} \frac{d^2(n)}{n} \\ &+ O\left(2^r \frac{q}{N} \max_{k|q} \left(\sum_{n \leq k/4} \frac{d^2(n)}{n \log k/n} \right)\right) + O\left(\frac{2^r}{N} \sum_{n \leq X} d^2(n)\right). \end{aligned}$$

The second error term is easily seen to be

$$\ll 2^r \frac{q}{N} T \log^3 X \ll 2^r T \log^3 X (\log \log 3q)^2.$$

We, now, prove

$$\begin{aligned} & \sum_{n \leq k/4} \frac{d^2(n)}{n \log k/n} \\ &\ll \log^3 k (\log \log 3k). \end{aligned}$$

This will give the first error term to be

$$\ll 2^r \frac{q}{N} \log^3 q (\log \log 3q) \ll 2^r \log^3 q (\log \log 3q)^3$$

writing $K = [\log(k/4)/\log 2]$, we have

$$\begin{aligned} & \sum_{n \leq k/4} \frac{d^2(n)}{n \log k/n} \\ &= \sum_{0 \leq m \leq K-1} \left(\sum_{2^m \leq n < 2^{m+1}} \frac{d^2(n)}{n \log k/n} \right) \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{0 \leq m \leq K-1} \frac{1}{2^m (\log k / 2^{m+1})} \left(\sum_{n \leq 2^{m+1}} d^2(n) \right) \\
&\ll \sum_{0 \leq m \leq K} \frac{m^3}{((\log k / \log 2) - m)} \\
&\ll \log^3 k (\log \log 3k).
\end{aligned}$$

We now prove

$$\begin{aligned}
&\sum_{\substack{n \leq X \\ (n, q)=1}} \frac{d^2(n)}{n} \\
&= \frac{1}{4\pi^2} \prod_{p|q} \left(\frac{1 - \frac{1}{p}}{1 - \frac{1}{p^2}} \right) \log^4 X + O(\log^3 X (\log \log 3q)^6),
\end{aligned}$$

consider

$$\frac{\zeta^4(s+1)}{\zeta(2s+2)} \eta(s) = \sum_{(n, q)=1} \frac{d^2(n)}{n^{s+1}},$$

$$\text{where } \eta(s) = \prod_{p|q} \left(\frac{1 - \frac{1}{p^{s+1}}}{1 - \frac{1}{p^{2s+2}}} \right).$$

We consider

$$\frac{1}{2\pi i} \int \frac{1 + (\log x)^{-1} + iX^s}{1 + (\log x)^{-1} - iX^s} \frac{\zeta^4(s+1)}{\zeta(2s+2)} \eta(s) \frac{X^s}{s} ds,$$

which is equal to

$$\sum_{\substack{n \leq X \\ (n, q)=1}} \frac{d^2(n)}{n} + O\left(\frac{X}{X^3} \sum_n \frac{d^2(n)}{n^{2+1/\log X} |\log(X/n)|}\right).$$

We now move the line of integration (in the above integral) to $\sigma = -1/4$. In doing so, we come across a pole at $s = 0$, the residue at which is

$$\frac{\eta(0) \log^4 X}{4\pi^2} + O(|a_1| \log^3 X),$$

where we write $\eta(s) = \eta(0) + a_1 s + a_2 s^2 + a_3 s^3 + \dots$

$$\text{Now } a_1 = \frac{1}{2\pi i} \int \frac{\eta(s)}{s^2} ds,$$

where we have taken the integral over the circle

$$|s| = \delta, \text{ with } \delta = \frac{1}{\log \log 3q}.$$

Thus $|a_1| \leq 1/\delta \max |\eta(s)|$, the maximum being taken over the circle $|s| = \delta$.

$$\text{Now } |\eta(s)| \ll \prod_{p|q} \left(1 + \frac{1}{p^{1-\delta}}\right)^4 \ll \exp\left(4 \sum_{p|q} \frac{1}{p^{1-\delta}}\right).$$

Now, we estimate

$$\begin{aligned} & \sum_{p|q} \frac{1}{p^{1-\delta}}; \\ & \sum_{p|q} \frac{1}{p^{1-\delta}} \leq \sum_{p \leq y} \frac{1}{p^{1-\delta}} + \sum_{\substack{p > y \\ p|q}} \frac{1}{p^{1-\delta}}. \end{aligned}$$

Further

$$\begin{aligned} & \sum_{p \leq y} \frac{1}{p^{1-\delta}} \leq y^\delta \sum_{p \leq y} \frac{1}{p} \\ & \leq y^\delta (\log \log y + A_1), \text{ where } A_1 \text{ is an absolute constant.} \end{aligned}$$

$$\begin{aligned} \text{And } \sum_{\substack{p > y \\ p|q}} \frac{1}{p^{1-\delta}} & \leq \sum_{\substack{p > y \\ p|q}} \frac{\log p}{y^{1-\delta} \log y} \\ & \leq \frac{1}{y^{1-\delta} \log y} \left(\sum_{p|q} \log p \right) \\ & \leq \frac{\log q}{y^{1-\delta} \log y}. \end{aligned}$$

$$\text{Thus } \sum_{p|q} \frac{1}{p^{1-\delta}} \leq y^\delta \left\{ \log \log y + A_1 + \frac{\log q}{y \log y} \right\}$$

choosing

$$y = \log 3q \text{ and since } \delta = \frac{1}{\log \log 3q}.$$

We get

$$\sum_{p|q} \frac{1}{p^{1-\delta}} \leq \log \log \log 3q + A_2$$

where A_2 is an absolute constant. Thus $|\eta(s)| \ll (\log \log 3q)^5$. Thus the residue at $s = 0$ is equal to

$$\frac{\eta(0) \log^4 X}{4\pi^2} + O(\log^3 X (\log \log 3q)^5).$$

$$\begin{aligned} \text{Thus } \sum_{\substack{n \leq X \\ (n, q) = 1}} \frac{d^2(n)}{n} &= \frac{1}{4\pi^2} \eta(0) \log^4 X + O(\log^3 X (\log \log 3q)^5) \\ &+ O\left(\frac{1}{X^2} \sum_n \frac{d^2(n)}{n^{2+1/\log X} |\log(X/n)|}\right) \\ &+ O\left(\int_{-1/4-4X^3}^{-1/4+4X^3} \frac{\zeta^4(s+1)}{\zeta(2s+2)} \eta(s) \frac{X^s}{s} ds\right). \end{aligned}$$

The second error term is easily $O(1)$. The third error term is

$$\ll X^{-1/4} \prod_{p|q} (1 + p^{-3/4})^4 \int_0^T \frac{|\zeta^4(3/4 + it)|}{|-1/4 + it|} dt.$$

$$\text{Now } \prod_{p|q} (1 + p^{-3/4})^4$$

$$\ll \exp(4 \sum_{p|q} p^{-3/4})$$

$$\ll \exp(A_3 (\log q)^{1/4} \log \log \log 3q), \text{ where } A_3 \text{ is an absolute constant.}$$

$$\text{Now } \int_0^T \frac{|\zeta^4(3/4 + it)|}{|-1/4 + it|} dt$$

$$\ll \sum_{m \geq 0} \int_{T/2^{m+1}}^{T/2^m} \frac{|\zeta^4(3/4 + it)|}{|-1/4 + it|} dt$$

$$\ll \sum_{m \geq 0} \frac{2^{m+1}}{T} \int_{T/2^{m+1}}^{T/2^m} |\zeta^4(3/4 + it)| dt$$

$$\ll (\log T)^5, \text{ by using mean fourth power of zeta function.}$$

Thus the third error term

$$\ll X^{-1/4} \log^5 T \cdot \exp(A_3 (\log q)^{1/4} \log \log \log 3q) \ll 1.$$

$$\text{Thus } \sum_{\substack{n \leq X \\ (n, q) = 1}} \frac{d^2(n)}{n} = \frac{\eta(0)}{4\pi^2} \log^4 X + O(\log^3 X (\log \log 3q)^5).$$

This completes the proof of the lemma.

Lemma 7. We have

$$\frac{1}{N} \sum_{\chi}^* \int_T^{2T} |J_k^2(s, \chi)| dt \ll 2^r \frac{q}{N} T (\log X)^3$$

for $3 \leq k \leq 6$.

Proof. To estimate the left hand side, we apply lemma 5 directly for $k = 3, 4$. For $k = 3$, while estimating, we note that $|1 - \exp(-n/X)| \ll n/X$, for $n \leq X$. The cases $k = 4, 6$ we sketch below. The case $k = 5$ can be disposed of, on similar lines.

We have

$$\begin{aligned} & \frac{1}{N} \sum_{\chi}^* \int_T^{2T} |J_k^2(s, \chi)| dt \\ & \ll X^{2\sigma-1} \left\{ 2^r \frac{q}{N} T \sum_{n > X} \frac{d^2(n)}{n^{2\sigma}} \exp(-2n/X) \right. \\ & \quad \left. + \frac{2^r}{N} \sum_{n > X} \frac{d^2(n)}{n^{2\sigma-1}} \exp(-2n/X) \right\}. \end{aligned}$$

We deal with the first, viz.,

$$X^{2\sigma-1} 2^r \frac{q}{N} T \sum_{n>X} \frac{d^2(n)}{n^{2\sigma}} \exp(-2n/X).$$

The second term can be similarly dealt with.

$$\text{In } \sum_{n>X} \frac{d^2(n)}{n^{2\sigma}} \exp(-2n/X),$$

by summing over intervals $[2^m, 2^{m+1})$ as done before, we get

$$\begin{aligned} \sum_{n>X} \frac{d^2(n)}{n^{2\sigma}} \exp(-2n/X) &\ll \sum_{m \geq M} 2^{m(1-2\sigma)} m^3 \exp(-2^{m+1} X^{-1}) \\ &\ll \sum_{r=0}^{\infty} 2^{(M+r)(1-2\sigma)} (M+r)^3 \exp(-2^{M+r+1} X^{-1}), \end{aligned}$$

where $M = [\log X / \log 2]$.

This, in turn, $\ll 2^{M(1-2\sigma)} M^3$.

$$\begin{aligned} &(\sum_{r=0}^{\infty} 2^{r(1-2\sigma)} (1+r/M)^3 \exp(-2^{M+r+1} X^{-1})) \\ &\ll 2^{M(1-2\sigma)} M^3 \ll X^{1-2\sigma} \log^3 X. \end{aligned}$$

Now we sketch the proof that

$$\frac{1}{N} \sum_{\chi}^* \int_T^{2T} |J_s^2(s, \chi)| dt \ll 2^r \frac{q}{N} T \log^3 X.$$

$$\begin{aligned} \text{The l.h.s.} &\ll X^{2\sigma-1} \int_T^{2T} dt \frac{1}{N} \sum_{\chi}^* \cdot \int_{u=1/4} |\psi^4(s+w, \chi)| \\ &\times \left| \sum_{n \leq X} d(n) \overline{\chi(n)} n^{w+s-1} \right|^2 X^{2u} |\Gamma(w)| |dw| \end{aligned}$$

by using Hölder's inequality. Breaking the integral with respect to ω at $|v| \leq \log^2 T$, where $v = \text{Im} \omega$, with a small error, we get the above with $u = 1/4$ to be

$$\begin{aligned} &\ll X^{1-2u-2\sigma} \int_{\substack{u=1/4 \\ |v| \leq \log^2 T}} |d\omega| |\Gamma(\omega)| \\ &\times \left(\int_T^{2T} dt \frac{1}{N} \sum_{\chi}^* \left| \sum_{n \leq X} d(n) \overline{\chi(n)} n^{\omega+s-1} \right|^2 \right). \end{aligned}$$

Using lemma 5, this expression

$$\ll X^{1-2u-2\sigma} \left\{ 2^r \frac{q}{N} \sum_{n \leq X} d^2(n) n^{2u+2\sigma-2} + \frac{2^r}{N} \sum_{n \leq X} d^2(n) n^{2u+2\sigma-1} \right\}.$$

The proof now can be completed.

Lemma 8 : We have

- (a) $\sum_{\chi}^* \int_T^{2T} |J_1^2(s, \chi)| dt \ll 2^r qT \log^3 X$ for $\sigma \leq \frac{1}{2} - 10^{-8}$ and
 (b) $\sum_{\chi}^* \int_T^{2T} |J_2^2(s, \chi)| dt \ll 2^r qT \log^3 X$ for $\sigma \geq \frac{1}{2} + 10^{-8}$.

Lemma 9. Let $s = \sigma + it$ with $t = T$ or $2T$. Then

- (a) $\int_{\frac{1}{2}-10^{-8}}^{\frac{1}{2}} d\sigma \sum_{\chi}^* |J_1^2(s, \chi)| \ll 2^r \cdot qT \log^3 X \cdot (\log \log 3q)$
 (b) $\int_{\frac{1}{2}}^{\frac{1}{2}+10^{-8}} d\sigma \sum_{\chi}^* |J_2^2(s, \chi)| \ll 2^r qT \cdot \log^3 X \cdot (\log \log 3q)$
 (c) $\sum_{\chi}^* |J_k^2(s, \chi)| \ll 2^r \cdot qT \log^3 X$
 for $\frac{1}{2} - 10^{-8} \leq \sigma \leq \frac{1}{2} + 10^{-8}$ and $3 \leq k \leq 6$.
 (d) $\int_{\frac{1}{2}-10^{-8}}^{\frac{1}{2}} d\sigma \sum_{\chi}^* |J_1(s, \chi) J_k(s, \chi)| \ll 2^r \cdot qT \log^3 X (\log \log 3q)$
 for $3 \leq k \leq 6$.
 (e) $\int_{\frac{1}{2}}^{\frac{1}{2}+10^{-8}} d\sigma \sum_{\chi}^* |J_2(s, \chi) J_k(s, \chi)| \ll 2^r \cdot qT \cdot \log^3 X \cdot (\log \log 3q)$
 for $3 \leq k \leq 6$.

Proof: We prove (a). The proof of (b) follows on the same lines. Similarly (c) can be proved. First, using Hölder's inequality with respect to integration (in σ) and then using (a), (b) and (c) the proof of (d) and (e) follows.

Now $\sum_{\chi}^* |J_1^2(s, \chi)|$

$$\begin{aligned} &\ll (qT)^{2\sigma-1} \sum_{\chi}^* \left| \sum_{n < X} \frac{d(n) \chi(n)}{n^s} \right|^2 \\ &\ll (qT)^{2\sigma-1} \sum_{m, n < X} \sum \frac{d(m) d(n)}{m^{\sigma+it} \cdot n^{\sigma-it}} \left(\sum_{\chi}^* \chi(m) \overline{\chi(n)} \right) \\ &\ll \sum_{k|q} |\mu(q/k)| \phi(k) X^{2\sigma-1} \sum_{m \equiv n \pmod{k}} \sum \frac{d(m) d(n)}{(mn)^{\sigma}} \\ &\ll \sum_{k|q} |\mu(q/k)| \phi(k) X^{2\sigma-1} \sum_{j=1}^k \left\{ \frac{d(j)}{j^{\sigma}} + \sum_{1 \leq n' \leq (X-j)/k} \frac{d(n'k+j)}{(n'k+j)^{\sigma}} \right\}^2 \\ &\ll \sum_{k|q} |\mu(q/k)| \phi(k) X^{2\sigma-1} \sum_{j=1}^k \left\{ \frac{d^2(j)}{j^{2\sigma}} + \left(\sum_{n' \geq 1} \frac{d(n'k+j)}{(n'k+j)^{\sigma}} \right)^2 \right\} \end{aligned}$$

Thus $\int_{\frac{1}{2}-10^{-8}}^{\frac{1}{2}} d\sigma \sum_{\chi}^* |J_1^2(s, \chi)|$

$$\ll \sum_{k|q} \frac{\phi(k) |\mu(q/k)|}{X} \left\{ \sum_{j=1}^k d^2(j) \int_{\frac{1}{2}-10^{-8}}^{\frac{1}{2}} \left(\frac{X}{j} \right)^{2\sigma} d\sigma \right.$$

$$\begin{aligned}
& + \sum_{j=1}^k \sum_{n' \geq 1} \left\{ \left(\frac{X}{n'k+j} \right)^{1/2} d(n'k+j) \right\}^2 \\
& \ll \sum_{k|q} \phi(k) |\mu(q/k)| \left\{ \sum_{j=1}^k \frac{d^2(j)}{j \log(X/j)} + \sum_{j=1}^k \left(\sum_{n' \geq 1} \frac{d^2(n'k+j)}{(n'k+j)^{1/2}} \right)^2 \right\}.
\end{aligned}$$

It can be shown easily that

$$\sum_{j=1}^k \frac{d^2(j)}{j \log(x/j)} \ll \log^3 k \cdot \log \log 3k$$

and using Hölder's inequality,

$$\begin{aligned}
& \sum_{j=1}^k \left(\sum_{n' \geq 1} \frac{d(n'k+j)}{(n'k+j)^{1/2}} \right)^2 \ll \left(\sum_{1 \leq n' \leq X/k} \frac{1}{(n'k)^{1/2}} \right) \\
& \times \left(\sum_j \sum_{1 \leq n' \leq (X-j)/k} \frac{d^2(n'k+j)}{(n'k+j)^{1/2}} \right) \ll (X^{1/2}k) \sum_{n \leq X} \frac{d^2(n)}{n^{1/2}} \ll X/k \log^3 x.
\end{aligned}$$

The proof can now be completed.

Lemma 10. We have

$$(a) \sum_X^* \int_T^{2T} J_i^2(\chi) dt \ll 2^r q T \log^3 X \cdot (\log \log 3q)$$

for $i = 1, 2$.

$$(b) \sum_X^* \int_T^{2T} J_i(\chi) J_k(\chi) dt \ll 2^r q T \log^3 X \cdot (\log \log 3q)$$

for $i = 1, 2$ and $3 \leq k \leq 6$.

Proof: We prove only

$$\sum_X^* \int_T^{2T} J_1^2(\chi) dt \ll 2^r q T \log^3 X \cdot (\log \log 3q).$$

To prove this, we move the line of integration from $\sigma = \frac{1}{2}$ to $\sigma = \frac{1}{2} - 10^{-8}$. Appealing to lemma 9 (a) and lemma 8 (a), the proof can be completed.

This completes the proof of the theorem.

Reference

- [1] Montgomery H L and Vaughan R C 1974 *J. London Math. Soc.* **8** 73-82

Note added in proof

Using

$$\begin{aligned}
& \int_T^{T+H} \left| \sum_n a(n) \chi(n) n^{-it} \right|^2 dt \\
& = H \sum_{\substack{n \\ (n,q)=1}} |a^2(n)| + O \left(\sum_n n |a^2(n)| \right)
\end{aligned}$$

(which can be trivially proved using Hilbert's inequality), our above proof can be trivially modified to give the result that 'for a primitive character $\chi \pmod{q}$ and for $T \geq 2$,

$$\int_T^{2T} \left| L\left(\frac{1}{2} + it, \chi\right) \right|^4 dt = 2T \sum_{\substack{n \leq qT \\ (n, q) = 1}} \frac{d^2(n)}{n} + O(qT \log^3 qT).$$

PROCEEDINGS (Mathematical Sciences)

Vol. 90, January - December 1981

SUBJECT INDEX (Mathematical Sciences)

- Abelian group
 - Polarisations on an abelian variety 125
- Balasubramanian's theorem
 - On the mean square value of Hurwitz zeta function 195
- Bifurcation diagrams
 - On some problems in singularity theory 1
- Boral summability
 - On the relation of generalized Valiron summability to Cesàro summability 181
- Boundary layers
 - Analyse asymptotique des equations de Transport dans le cas dévolution 219
- Bounded argument rotation bounded boundary rotation
 - On a generalization of the class of functions with bounded Mocanu variation 213
- Bundles
 - Yang-Mills and bundles over algebraic curves 11
- Carton subgroup
 - A submersion principle and its applications 95
- Chern-Gauss-Bonnet theorem
 - Curvature and the heat equation for the de Rham complex 47
- Clifford bundles
 - Orthogonal and spin bundles over hyper-elliptic curves 151
- Closure properties
 - EOL and ETOL array languages 167
- Cohomological computations
 - Yang-Mills and bundles over algebraic curves 11
- Complex singularity
 - On some problems in singularity theory 1
- Continuous kernels
 - On operators of trace class in $L^2(X, \mu)$ 29
- Convolution
 - On a generalization of the class of functions with bounded Mocanu variation 213
- Correctors
 - Homogenization of eigenvalue problems in perforated domains 239
- Curvature
 - Curvature and the heat equation for the de Rham complex 47
- de Rham complex
 - Curvature and the heat equation for de Rham complex 47
- Diffeomorphism
 - Deformations of metrics and associated harmonic maps 33
- Diffusion approximation
 - Analyse asymptotique des équations de Transport dans le cas d'évolution 219
- Dirichlet eigenvalue problem
 - Homogenization of eigenvalue problem in perforated domains 239
- Dirichlet problem
 - Deformations of metrics and associated harmonic maps 33
- Dirichlet series
 - Poisson formulae of Hecke type 129
- Discrete uniform subgroups
 - Geometric construction of cohomology for arithmetic groups I 103
- Distribution kernel
 - On operators of trace class in $L^2(X, \mu)$ 29
- Eigenvalues
 - Homogenization of eigenvalue problems in perforated domains 239
- Eigenvectors
 - Homogenization of eigenvalue problems in perforated domains 239
- Elliptic semilinear equations
 - Application of Newton's method to a homogenization problem 229
- EOL array systems
 - EOL and ETOL array languages 167
- ETOL array systems
 - EOL and ETOL array languages 167
- Fredholm alternative
 - Analyse asymptotique des équations de Transport dans le cas d'évolution 219

- Galois cocycle
 Geometric construction of cohomology for arithmetic groups I 103
- Generalized Valiron summability
 On the relation of generalized Valiron summability to Cesàro summability 181
- Geodesics
 Deformations of metrics and associated harmonic maps 33
- Harder-Narasimhan method
 Yang-Mills and bundles over algebraic curves 11
- Harmonic forms
 Vanishing theorems for square-integrable harmonic forms 21
- Harmonic maps
 Deformations of metrics and associated harmonic maps 33
- Heat equation
 Curvature and the heat equation for the de Rham complex 47
- Hierarchy
 EOL and ETOL array languages 167
- Hilbert-Schmidt operators
 On operators of trace class $L^2(X, \mu)$ 29
- Hill's operator
 Units of Hill curves 81
- Homogenization
 Application of Newton's method to a homogenization problem 229
- Homogenization
 Homogenization of eigenvalue problems in perforated domains 239
- Homomorphism
 Polarisations on an abelian variety 125
- Hybrid analogue for L-series
 On the mean square value of Hurwitz zeta function 195
- Hyperelliptic curves
 Orthogonal and spin bundles over hyperelliptic curves 151
- Hypergeometric function
 On a generalization of the class of functions with bounded Mocanu variation 213
- Integral operators
 On operators of trace class in $L^2(X, \mu)$ 29
- Invariance theory
 Curvature and the heat equation for the de Rham complex 47
- Killing degeneracy
 Deformations of metrics and associated harmonic maps 33
- Laplace operator
 Units of Hill curves 81
- Lie groups
 Geometric construction of cohomology for arithmetic groups I 103
- L-series
 A note on the mean value of L-series 273
- Matsushima's theorem
 Vanishing theorems for square-integrable harmonic forms 21
- Maxima singularities
 On some problems in singularity theory 1
- Mean value
 A note on the mean value of L-series 273
- Mellian transforms
 Poisson formulae of Hecke type 129
- Mocanu variation
 On a generalization of the class of functions with bounded Mocanu variation 213
- Modality
 On some problems in singularity theory 1
- Morse index
 Yang-Mills and bundles over algebraic curves 11
- Neron-Severi group
 Polarisations on an abelian variety 125
- Neumann eigenvalue problem
 Homogenization of eigenvalue problems in perforated domains 239
- Newton's method
 Application of Newton's method to a homogenization problem 229
- Non-negative Jacobian
 Deformations of metrics and associated harmonic maps 33
- Non-zero cohomology
 Geometric construction of cohomology for arithmetic groups I 103
- Normal form
 EOL and ETOL array languages 167
- Orthogonal bundles
 Orthogonal and spin bundles over hyperelliptic curves 151
- Orthogonal decomposition
 Geometric construction of cohomology for arithmetic groups I 103
- Patodi's cancellation lemma
 Curvature and the heat equation for the de Rham complex 47
- Poisson formulae
 Poisson formulae of Hecke type 129
- Polarisation
 Polarisations on an abelian variety 125
- Primitive characters
 A note on the mean value of L-series 273

- Rajagopal's theorem
 On the relation of generalized Valiron summability to Cesàro summability 181
- Ramanujan's function
 Poisson formulae of Hecke type 129
- Real singularity
 On some problems in singularity theory 1
- Riemannian metrics
 Deformations of metrics and associated harmonic maps 33
- Riemannian manifold
 Vanishing theorems for square-integrable harmonic forms 21
- Riemannian manifold
 Curvature and the heat equation for the de Rham complex 47
- Riemann surfaces
 Yang-Mills and bundles over algebraic curves 11
- Rosati involution
 Polarizations on an abelian variety 125
- Schwartz function
 Poisson formulae of Hecke type 129
- Schwartz kernel theorem
 On operators of trace class in $L^2(X, \mu)$ 29
- Singularity theory
 On some problems in singularity theory 1
- Spin bundles
 Orthogonal and spin bundles over hyper-elliptic curves 151
- Spin structure
 Orthogonal and spin bundles over hyper-elliptic curves 151
- Stekloff eigenvalue problem
 Homogenization of eigenvalue problems in perforated domains 239
- Submersion principle
 A submersion principle and its applications 95
- Synchronized form
 EOL and ETOL array languages 167
- Toda units
 Units of Hill curves 81
- Trace class operators
 On operators of trace class in $L^2(X, \mu)$ 29
- Trace formula
 Units of Hill curves 81
- Transport equations
 Analyse asymptotique des équations de Transport dans le cas d'évolution 219
- Unimodular group
 A submersion principle and its applications 95
- Vanishing theorems
 Vanishing theorems for square-integrable harmonic forms 21
- Yang-Mills equations
 Yang-Mills and bundles over algebraic curves 11

AUTHOR INDEX (Mathematical Sciences)

- Arnold V
 On some problems in singularity theory 1
- Atiyah M F
 Yang-Mills and bundles over algebraic curves 11
- Bott R
 see Atiyah M F 11
- Dodziuk Jozef
 Vanishing theorems for square-integrable harmonic forms 21
- Duistermaat J J
 On operators of trace class in $L^2(X, \mu)$ 29
- Eells J
 Deformations of metrics and associated harmonic maps 23
- Gilkey Peter B
 Curvature and the heat equation for the de Rham complex 47
- Harish-Chandra
 A submersion principle and its applications 95
- Kamala Krithivasan
 see Nalinakshi Nirmal 167
- Kesavan S
 Application of Newton's method to a homogenization problem 229
- Krishnan V K
 On the relation of generalized Valiron summability to Cesàro summability 181
- Lemaire L
 see Eells J 33

- | | | | |
|--|-----|---|-----|
| McKean H P | | Ramanan S | |
| Units of Hill curves | 81 | Orthogonal and spin bundles over hyper-elliptic curves | 151 |
| Millson John J | | Rane V V | |
| Geometric construction of cohomology for arithmetic groups I | 103 | A note on the mean value of L -series | 273 |
| Nalinakshi Nirmal | | Rangachari S S | |
| EOL and ETOL array languages | 167 | see Raghavan S | 129 |
| Narasimhan M S | | Santos Rafael F | |
| Polarisations on an abelian variety | 125 | Analyse asymptotique des équations de Transport dans le cas d'évolution | 219 |
| Narlikar M J | | Thangamani J | |
| On the mean square value of Hurwitz zeta function | 195 | On a generalization of the class of functions with bounded Mocanu variation | 213 |
| Nori M V | | | |
| see Narasimhan M S | 125 | | |
| Raghavan S | | Vanninathan M | |
| Poisson formulae of Hecke type | 129 | Homogenization of eigenvalue problems in perforated domains | 239 |
| Raghunathan M S | | | |
| see Millson John J | 103 | | |

